

MESTRADO EM CIÊNCIA DA INFORMAÇÃO

# **GESTÃO DE DADOS DE INVESTIGAÇÃO DO TIPO EXPERIMENTAL: casos de uso e contribuições para a melhoria da qualidade dos metadados**

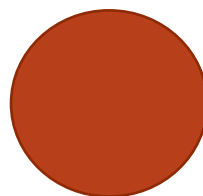
Cristiana Sofia Pereira Landeira

**M**

**2018**

UNIDADES ORGÂNICAS ENVOLVIDAS

FACULDADE DE ENGENHARIA  
FACULDADE DE LETRAS



CRISTIANA SOFIA PEREIRA LANDEIRA

**Gestão de dados de investigação do tipo experimental:  
casos de uso e contribuições para a melhoria da  
qualidade dos metadados**

Dissertação realizada no âmbito do Mestrado em Ciência da Informação

Orientador: Prof. Dra. Maria Cristina de Carvalho Alves Ribeiro

Coorientador: João Daniel Aguiar Castro

Faculdade de Engenharia da Universidade do Porto

Junho de 2018

CRISTIANA SOFIA PEREIRA LANDEIRA

**Gestão de dados de investigação do tipo experimental:  
casos de uso e contribuições para a melhoria da  
qualidade dos metadados**

Dissertação realizada no âmbito do Mestrado em Ciência da Informação

Orientador: Prof. Dra. Maria Cristina de Carvalho Alves Ribeiro

Coorientador: João Daniel Aguiar Castro

**Membros do Júri**

Presidente: Professor Doutor António Manuel Lucas Soares

Faculdade de Engenharia da Universidade do Porto

Arguente: Professora Doutora Mariana Curado Malta

Instituto Superior de Contabilidade e Administração do Porto

Orientador: Professora Doutora Maria de Carvalho Alves Ribeiro

Faculdade de Engenharia da Universidade do Porto

## Agradecimentos

A caminhada foi longa, cheia de desafios, sem dúvida a mais difícil até agora e foram vários que contribuíram para que ela se realizasse, sem os quais não seria possível, e por essa razão devo um agradecimento sincero.

Quero agradecer à minha orientadora Prof. Doutora Cristina Ribeiro, por me acolher neste projeto de dissertação, pela disponibilidade que sempre mostrou e pelo exemplo de profissional a seguir.

Ao meu coorientador João Castro por toda a disponibilidade, ajuda e pela partilha de conhecimento.

A todos os investigadores que disponibilizaram do seu tempo para colaborar na avaliação dos casos de uso sem eles não seria possível obter os mesmos resultados.

A toda a equipa de investigadores do InfoLab por todos os momentos que tornaram os dias mais fáceis.

Aos meus pais, e irmão que apesar das dificuldades sempre me incentivaram a estudar, agradeço todo o esforço e dedicação, agradeço todo o carinho e segurança que sinto convosco.

À minha Mariazinha e ao meu Paulino que tornaram a minha vida tão especial, tanto me fizeram rir, que me deram um pedacinho de cada um, infelizmente não tivemos tempo de festejar esta vitória juntos, mas sinto que me continuam a dar a mão, como quando precisava de ajuda para saltar o rio.

A toda a minha família e amigos, cada um com o seu jeito especial, me mostram todos os dias que sou uma pessoa cheia de sorte por ter pessoas tão especiais do meu lado.

À colega de casa por todas as conversas, risadas, pelas cevadas que interrompiam ciclos de trabalho, sabes a importância que tens!

Ao meu namorado, pela calma, força e suporte que sempre me soubeste dar em todos os momentos.

A todos um obrigada muito especial por fazerem parte da concretização de um ciclo muito importante na minha vida, repleto de oportunidades e de sonhos concretizados.

## Resumo

O atual contexto de produção de dados de investigação leva ao desenvolvimento de ferramentas para os desafios de gestão. Os metadados sustentam a partilha dos dados, amplamente defendida pelas agências de financiamento. Contudo o desenvolvimento de modelos de metadados nem sempre se apresenta uma tarefa fácil, uma vez que é necessário conciliar a colaboração do curador de dados e do investigador. Cada um deles tem um conhecimento diferente sobre os domínios e a gestão de dados, tornando a comunicação um processo exigente.

Este trabalho tem como objetivo compreender se o curador, sem conhecimento específico sobre um determinado domínio, consegue definir um modelo de metadados preliminar baseado na análise de conteúdo de artigos produzidos por investigadores. Pretende-se compreender se a abordagem permite otimizar o fluxo de trabalho do curador e ainda se é possível definir requisitos que se relacionem com a tipologia de dados, neste caso em específico os dados experimentais. De forma a avaliar esta abordagem foram efetuadas análises de conteúdo em publicações científicas para cinco casos de uso para os quais resultam modelos de metadados que são avaliados junto de investigadores do domínio. Através da avaliação junto dos investigadores é possível validar a abordagem com resultados satisfatórios, uma vez que a taxa de conceitos aceite foi elevada. Na perspetiva do curador de dados a análise de conteúdo tem potencial como uma tarefa proativa que pode ser aplicada em diversos domínios experimentais e possibilita uma comunicação mais eficiente entre o curador e os investigadores.

No entanto, a abordagem também tem algumas limitações que se relacionam com a falta de conhecimento do curador sobre as especificidades dos domínios. A interação final entre curadores e investigadores é essencial no desenvolvimento do modelo de metadados.

**PALAVRAS-CHAVE:** Análise de conteúdo; gestão de dados de investigação; metadados; energias renováveis; engenharia mecânica; física; química sustentável

## Abstract

The current context of research data production leads to the development of tools to face the challenges of the management and the description of data that result from the research processes. Metadata support data sharing, widely advocated by funding agencies. However, the development of metadata is not always an easy task, since it is necessary to promote the collaboration of the data curator and the researcher. They have complementary knowledge on domains and data management, making communication a demanding process.

The objective of this work is to understand if the curator, without specific knowledge about the domain, can define a preliminary data model, using content analysis of papers by the researchers. This approach is expected to optimize the workflow of the curator and to test the hypothesis that it is possible to define requirements that are related to the typology of data, in this specific case experimental data. In order to examine this approach, content analyses were carried out based on scientific publications for five use cases, for which metadata models are produced and evaluated with domain researchers. Evaluation with the researchers has provided evidence to support the approach. The results are promising with a high rate of accepted concepts. In the data curator's perspective, content analysis has potential as a proactive task that can be applied in several experimental domains and enables a more efficient communication between the curator and the researchers.

However, the approach also has some limitations that relate to the curator's lack of knowledge about the specificities of the domains, so that the final interaction between curators and researchers is essential to bridge some of the gaps in the model.

**KEYWORDS:** Content analysis; research data management; metadata; renewable energy; mechanical Engineering; physical; sustainable chemistry

# Sumário

Agradecimentos	4
Resumo	5
Abstract	6
Sumário	7
Lista de Figuras	9
Lista de Tabelas	10
Siglas e abreviaturas	11
I Introdução	12
1.1 Motivação e definição de objetivos	13
1.2 Estrutura da dissertação	14
2 Gestão de Dados de Investigação	17
2.1 Dados de investigação	21
2.2 Plataformas de Gestão de dados de investigação	26
2.2.1 Zenodo	28
2.2.2 Figshare	29
2.2.3 Dryad	30
2.2.4 CKAN	31
2.2.5 B2SHARE	32
2.2.6 Re3Data	33
2.3 Metadados para a descrição de dados de investigação	33
2.3.1 DC (Dublin Core)	37
2.3.2 DDI (Data Documentation Initiative)	38
2.3.3 EML (Ecological Metadata Language)	39
2.3.4 DIF (Directory Interchange Format)	39
2.3.5 CIF (Crystallographic Information Framework)	40
2.4 Sumário	41
3 Metodologia	42
4 Casos de Uso	44
4.1 Química sustentável - Degradação de partículas poluentes	45
4.2 Energias Renováveis - Produção de camadas absorventes – Filmes finos	48
4.3 Física - Síntese de nanopartículas	52
4.4 Engenharia Mecânica - Mecânica de fratura: Feixe cantiléver	56
4.5 Engenharia Mecânica - Processo de fabrico: Moldagem de chapa	60
5 Discussão	64

6	Conclusões	68
6.1	Limitações do estudo	69
6.2	Investigação futura	70
7	Referências Bibliográficas	71
Anexo I. Tabela de análise do Caso de Uso Química sustentável - Remoção de partículas poluentes		74
Anexo II. Tabela de resultados do Caso de Uso Química sustentável- Remoção de partículas poluentes		79
Anexo III. Tabela de análise do Caso de Uso Energias Renováveis - Produção de camadas absorventes: Filmes finos		84
Anexo IV. Tabela de resultados do Caso de Uso: Energias Renováveis - Produção de camadas absorventes - Filmes Finos		89
Anexo V. Tabela de análise do Caso de Uso da Física - Síntese de nanopartículas		92
Anexo VI. Tabela de resultados do Caso de Uso da Física - Síntese de nanopartículas		99
Anexo VII. Tabela de análise do Caso de Uso Engenharia Mecânica - Mecânica de fratura: Feixe Cantiléver		105
Anexo VIII. Tabela de resultados do Caso de Uso Engenharia Mecânica - Mecânica de fratura - Feixe Cantiléver		110
Anexo IX. Tabela de análise do Caso de Uso Engenharia Mecânica - Processo de fabricação - Moldagem de chapa		113



## Lista de Figuras

Figura 1-Classificação das áreas de investigação segundo FORD	24
Figura 2-Classificações de dados Kotani (1975)	25
Figura 3-Metadados Zenodo	29
Figura 4-Metadados Zenodo	29
Figura 5-Metadados Figshare	30
Figura 6-Metadados Dryad	31
Figura 7-Metadados CKAN	32
Figura 8-Metadados B2SHARE	32
Figura 9-Requisitos fundamentais para a descrição de dados de investigação	36
Figura 10-Processo que origina os casos de uso	44
Figura 11-Descritores identificados pelo curador	57

## Lista de Tabelas

Tabela 1-Resumo de resultados obtidos para a química sustentável: degradação de partículas poluentes .....	45
Tabela 2-Requisitos necessários para a descrição da química sustentável: degradação de partículas poluentes .....	47
Tabela 3-Resumo de resultados obtidos para Energias Renováveis: Produção de camadas absorventes: filmes finos .....	49
Tabela 4-Requisitos necessários para a descrição das Energias renováveis: Produção de Camadas Absorventes: filmes finos .....	50
Tabela 5-Resumo de resultados obtidos para Física-Síntese de nanopartículas.....	53
Tabela 6-Resumo de resultados obtidos para Física - Síntese de nanopartículas (descritores ambíguos).....	55
Tabela 7-Requisitos necessários para a descrição da Física - Síntese de nanopartículas .....	55
Tabela 8-Resumo de resultados obtidos para Engenharia Mecânica - Mecânica de fratura: feixe cantiléver .....	58
Tabela 9-Requisitos necessários para a descrição de Engenharia Mecânica – Processo de fabrico: moldagem de chapa.....	62
Tabela 10-Resumo dos aspetos positivos e negativos da análise de conteúdo .....	68

## Siglas e abreviaturas

BODC	British Oceanographic Data Centre
CERN	European Organization for Nuclear Research
CIF	Crystallographic Information Framework
CKAN	Comprehensive Knowledge Archive Network
CSD	Cambridge Structural Database
DataONE	Data Observation Network for Earth
DC	Dublin Core
DCC	Digital Curation Centre
DDI	Data Documentation Initiative
DIF	Directory Interchange Format
DOI	Digital Object Identifier
EML	Ecological Metadata Language
EPSRC	Engineering and Physical Sciences Research Council
ESADS	Earth Science and Applications Data Systems
FAIR	Findable, Accessible, Interoperable and Reusable
FEUP	Faculdade de Engenharia da Universidade do Porto
GBIF	Global Biodiversity Information Facility
IUCr	International Union of Crystallography
JISC	Joint Information Systems Committee
MANTRA	Research Data Management Training
NSF	National Science Foundation
OECD	Organisation for Economic Co-operation and Development
RCAAP	Repositórios Científicos de Acesso Aberto de Portugal
RDA	Research Data Alliance
RDM	Research Data Management
XML	Extensible Markup Language

# I Introdução

Num contexto em que a valorização da partilha de dados de investigação é crescente, onde se adota o conceito da “ciência aberta” e se fomenta o trabalho colaborativo entre várias áreas científicas, é necessário disponibilizar às comunidades científicas soluções ajustadas às necessidades dos mais diversos domínios. Apesar da evolução tecnológica e do aumento de investimento no desenvolvimento de soluções é notória a falta de ferramentas e modelos que permitam a gestão de dados em comunidades científicas específicas.

Enquanto que para as *big sciences*, como a astronomia e a física, se desenvolveram ferramentas ajustadas às necessidades de descrição dos dados de investigação, para as comunidades da “cauda longa” (Heidorn 2008) existe uma carência de infraestruturas para a gestão dos dados de investigação. Assim, um dos grandes desafios colocados à gestão de dados é o desenvolvimento de metadados, atividade fundamental que garante a recuperação, acesso e interpretação dos dados.

São vários os benefícios apontados à partilha dos dados de investigação, como a melhoria da qualidade dos resultados influenciada pelo aumento da transparência que é garantida pela consulta dos dados por outros investigadores. O incentivo à investigação colaborativa, que diminui a duplicação de esforços e acelera a inovação (European Commission 2017). Ou ainda a validação de métodos de investigação bem como o reconhecimento do investigador gerado através do aumento das citações (UK Data Archive 2011). Contudo, os benefícios não são óbvios para quem produz os dados, possivelmente pela quantidade de recursos necessários para a sua gestão (Castro 2013). O tratamento de dados exige muito esforço, desenvolvendo-se à custa de conhecimento especializado, ou ainda pela adoção lenta de ferramentas e serviços que dificulta o processo (Wallis, Rolando, e Borgman 2013; RCAAP 2010). A reutilização dos dados de investigação está dependente do enriquecimento semântico (Thanos, 2017), ou seja, é necessário adicionar aos dados de investigação metadados suficientes que garantam a sua recuperação, interpretação e acesso a longo prazo.

A interação entre os vários stakeholders envolvidos na gestão de dados de investigação é preponderante para o desenvolvimento de ferramentas ajustadas às necessidades dos domínios. Em especial a colaboração entre os investigadores e os curadores de dados o primeiro porque possui conhecimento específico sobre o domínio (White 2014) e o segundo porque detém o conhecimento de gestão da informação. No entanto, esta interação nem sempre é possível, por um lado os investigadores não têm uma cultura de partilha de dados de investigação e por outro, o processo de investigação apresenta-se como prioridade. Considerando ainda que as dificuldades que as pequenas comunidades enfrentam pela falta de ferramentas, ou de financiamento, explicam a relutância na colaboração para a gestão de dados, da qual não retiram benefícios diretos.

A necessidade de distinguir recursos leva ao desenvolvimento de várias classificações, assim os dados de investigação podem ser categorizados de acordo com a origem; nível de tratamento; área científica que os produziu ou utilizou, entre outras. A diversidade de classificações de dados de investigação e as suas origens contextuais sugerem que os metadados, desenvolvidos para documentar os dados de investigação, terão características diferentes que se relacionam com cada tipologia de dados. Isso irá influenciar os requisitos a considerar na produção de metadados.

O contributo deste projeto assenta na necessidade de identificar requisitos de metadados que possam assegurar que os dados científicos são bem descritos, garantindo assim a possibilidade de serem recuperados, interpretados e utilizados a longo prazo.

Pretende-se aumentar o conhecimento relativo à descrição de dados em contexto experimental sendo para isso considerada uma abordagem baseada na análise de documentos produzidos por investigadores que, tal como Chao (2014) defende, é uma alternativa viável que possibilita a extração de informação relevante para o desenvolvimento de metadados. A análise de conteúdos foi aplicada em cinco casos de uso, um do domínio da química sustentável; outro da energia renovável, dois casos da engenharia mecânica e um da física. A seleção dos casos foi orientada de acordo com relações anteriormente estabelecidas com os investigadores, dada uma maior facilidade de interação com os mesmos. O foco recai apenas sobre os dados experimentais uma vez que, para os dados observacionais (registos de produção dispendiosa e de difícil ou impossível reprodução) são disponibilizadas ferramentas em maior número do que nos outros casos. Por sua vez, os dados provenientes de contextos experimentais e simulacionais, em que é possível a reprodução se existir informação relativa às variáveis, metodologias de recolha e configuração das experiências, não existem muitas soluções. Para estes dois casos, o suporte à descrição dos dados de investigação é escasso, o que inibe a produção de metadados, impossibilitando a reprodução, acesso, reutilização e ainda na maioria dos casos leva à perda irreversível dos dados.

## **1.1 Motivação e definição de objetivos**

Acredita-se que com o aumento do conhecimento relativo aos requisitos necessários para a descrição dos dados é possível auxiliar o trabalho desenvolvido pelos investigadores na produção de metadados. Os metadados permitem descrições completas sobre os conjuntos de dados, possibilitando a sua recuperação acesso e difusão. Fomentando um ambiente multidisciplinar científico, através da partilha dos dados de investigação. Na chamada “cauda longa” começam a surgir ferramentas e métodos para a gestão de dados, assim como repositórios e vários modelos de metadados. Quando bem documentados os dados experimentais podem ser reproduzíveis, contudo não estão disponíveis metadados suficientes

que captem a informação necessária que permita essa reprodução. Nesse sentido, o foco deste trabalho são os dados do tipo experimental uma vez que para os dados observacionais (registros resultantes de observações diretas e que na maioria dos casos não podem ser reproduzidos pela dependência do tempo e espaço), para estes casos já existem soluções bem definidas.

O desenvolvimento deste projeto de dissertação permite aumentar o conhecimento relativo à descrição de dados do tipo experimental. Assim, através da análise de casos de estudo, pretende-se identificar requisitos de metadados que ajudarão a suprir necessidades específicas dos domínios.

As oportunidades que motivam o desenvolvimento deste projeto são:

- a legislação do acesso aberto, que propõe que os resultados de investigação científica devem ser partilhados, para que possam ser reutilizados por outros e assim garantir o desenvolvimento do conhecimento científico;
- o pouco conhecimento relativo aos modelos de dados necessários para a descrição de dados do tipo experimental;
- o aumento do interesse por parte das agências de financiamento em estabelecer regulamentos que adotam a visão da ciência aberta, particularmente nos casos em que as investigações são financiadas por fundos públicos.

O principal objetivo deste trabalho é verificar se o curador de dados pode autonomamente, sem conhecimento prévio em domínios específicos, compreender e identificar conceitos necessários à descrição de dados nos mais diversos contextos de investigação, sobretudo naqueles onde não estão disponíveis ferramentas para o efeito. Não se pretende com isto substituir a interação com o investigador, considerando-se uma fase preponderante. Contudo acredita-se que a interação pode ser mais rentável se o curador de dados tiver identificado um modelo preliminar, e assim facilitar a interação entre o curador de dados e o investigador. Sendo que o último deve validar o modelo de metadados e sugerir requisitos complementares, para a qualidade dos metadados. Uma vez que, o curador por si só, poderá não ter conhecimento para validar metadados específicos dos domínios.

## **1.2 Estrutura da dissertação**

Esta dissertação está organizada da seguinte forma:

A revisão da literatura aborda quatro tópicos principais que são: gestão de dados de investigação; dados de investigação; plataformas de gestão de dados de investigação e por fim, metadados para a descrição de dados de investigação. Esta abordagem permite efetuar uma contextualização sobre questões fundamentais que darão suporte às etapas posteriores.

No primeiro tópico, é abordada a questão do aumento da produção dos dados de investigação, que pode não ocorrer por completo, uma vez que, para as pequenas comunidades científicas os recursos de gestão de dados são escassos. Num contexto em que cada vez mais se valoriza a ciência aberta a gestão dos dados de investigação apresenta-se como preponderante. Contudo o acesso, interpretação e reutilização dos dados de investigação está dependente do desenvolvimento de metadados que possam ser usados para a descrição dos conjuntos de dados. Um dos grandes desafios neste contexto é o desenvolvimento de metadados. Neste ponto são ainda analisados os benefícios e as barreiras da partilha de dados de investigação.

No segundo tópico, apresentam-se as definições para dados de investigação e as várias classificações que estes podem assumir. Assim, dados de investigação são registos factuais que podem ser números, imagens, sons ou registos textuais, usados como fonte para apoiar os resultados das investigações científicas. As várias classificações para os dados de investigação e as suas origens contextuais sugerem que os metadados usados para documentar o conteúdo, a estrutura e o contexto terão características distintas que se relacionam com cada tipologia de dados. Assim sendo, este fator irá influenciar os requisitos.

No terceiro tópico, constata-se que existem infraestruturas para a gestão de dados de investigação de comunidades bem estabelecidas, no entanto o mesmo não se verifica para comunidades pequenas. Reforça-se a importância do desenvolvimento de infraestruturas capazes de gerir os dados de investigação, só assim será possível a sua partilha, compreensão e reutilização a longo prazo. Contudo, no caso dos repositórios, o conjunto de metadados utilizados para a descrição dos dados é reduzido.

Assim, no último tópico da revisão de literatura são abordadas questões relacionadas com os metadados. Os metadados são fundamentais para garantir a autenticidade dos dados, atribuindo informação que permite aumentar a inteligibilidade garantindo assim o acesso, reutilização e preservação a longo prazo. Por fim, é efetuada a apresentação de alguns modelos de metadados como por exemplo a Data Documentation Initiative. Verifica-se que existem modelos de dados bem desenvolvidos para comunidades bem estabelecidas como é o caso do CIF para a cristalografia. Contudo as soluções mais gerais, como é o caso da Dublin Core, por si só não são suficientes para descrever domínios diversos, por essa razão é fundamental aumentar o conhecimento relativo aos requisitos necessários para os metadados nos contextos experimentais.

Na secção três apresenta-se a metodologia em que se baseia a análise de conteúdos produzidos por investigadores, com o objetivo de recolher informação que possibilite o desenvolvimento do modelo de metadados. O resultado da análise é apresentado sempre que possível a investigadores que procedem com a avaliação. Notando-se uma diferenciação na

interação com os investigadores numa tentativa de validar, de ajustar a abordagem à sensibilidade do investigador para a gestão de dados de investigação.

Na secção quatro são apresentados os resultados obtidos através da aplicação da metodologia aos domínios da química sustentável; energia renovável; física; engenharia mecânica – (1) feixe cantilever; (2) moldagem de chapa. Para cada um dos casos de uso é efetuada uma breve descrição do domínio seguindo-se a apresentação dos resultados obtidos da análise, bem como da validação dos modelos de dados.

Por fim, apresentam-se as secções relativas à discussão, onde serão discutidos alguns dos resultados obtidos, seguem-se as conclusões, limitações do estudo e investigação futura.



## 2 Gestão de Dados de Investigação

Atualmente são produzidas grandes quantidades de dados, resultando do aumento das ferramentas tecnológicas que apoiam a recolha, a produção, análise, tratamento e armazenamento de dados. No entanto, Borgman, Wallis e Mayernik (2007) defendem que o dilúvio de dados está a ocorrer, mas não por completo. Se por um lado as *big sciences*, desenvolveram ferramentas e repositórios capazes de lidar com o aumento da produção científica, por outro lado as pequenas ciências, devido à falta de recursos, necessitam de infraestruturas eficientes para gerir os dados.

A gestão dos dados de investigação é preponderante uma vez que os dados são recursos valiosos que exigem muito dinheiro e tempo para serem produzidos. A replicação de esforços para gerar dados anteriormente tratados seria minimizada se a partilha e reutilização fosse uma das preocupações da comunidade científica, das organizações de investigação, dos financiadores e dos curadores de dados. É notório que cada vez mais as agências de financiamento estabelecerem regulamentos que segundo a European Open Science Cloud (2016) adotam “*a visão da ciência aberta, em particular para a ciência financiada com fundos públicos, as agências de financiamento estão cada vez mais a requerer que os resultados de investigação, incluindo os dados, estejam disponíveis publicamente para outros grupos os usarem*” (Ribeiro et al. 2016).

Borgman (2012) defende que os dados de investigação assumem diversas formas, são tratados de vários modos, assumindo diversas abordagens, sendo com frequência difíceis de interpretar quando removidos do contexto que lhes deu origem, transformando a partilha de dados num enigma. Nesse sentido é preponderante que os dados sejam acompanhados de informação capaz de os contextualizar, possibilitando a partilha que se torna possível através do “enriquecimento do ponto de vista semântico” (Thanos 2017). Ou seja, a partilha de dados de investigação está dependente do desenvolvimento de estruturas de metadados que os investigadores possam usar para descrever os conjuntos de dados que geram.

Entre os vários desafios colocados à gestão de dados desde manter, preservar, de modo a permitir que estes sejam compreendidos por terceiros e possíveis de usar a longo prazo, a produção de metadados é uma atividade fundamental para que os dados possam ser recuperados, acedidos e interpretados. De acordo com o RCAAP (2010) a gestão de dados científicos engloba as ações de representação, armazenamento, associação de metadados (que descrevem, documentam e ajudam à compreensão dos mesmos) a organização dos dados em coleções, a indexação para pesquisa futura e todas as formas de apresentação dos dados.

Ribeiro et al. (2016) defendem que a gestão dos dados não está estabelecida a par do desenvolvimento dos projetos de investigação. E ainda que a gestão dos dados em áreas como as ciências da vida é um processo pensado e desenvolvido desde o início, ao contrário dos pequenos grupos de investigação que não têm tecnologia para gerir os dados, mas que têm um

grande potencial para gerar dados valiosos. Também Cisar et al. (2016) afirmam que existe, na maioria dos sistemas de gestão de dados, uma falha na infraestrutura para o uso diário do investigador.

Contudo, verifica-se que existem algumas instituições que promovem a gestão e a partilha de dados de investigação, como é o caso do “*NCBI nas ciências da vida e o ICPSR nas ciências sociais (...) infraestruturas que suportam as tarefas de curadoria de dados de comunidades bem estabelecidas*” (Ribeiro et al. 2016) Outro exemplo é a Research Data Alliance<sup>1</sup> (RDA) uma organização desenvolvida em 2013 por várias comunidades entre elas, a Comissão Europeia, a United States National Science Foundation, o National Institute of Standards and Technology, e o Australian Government’s Department of Innovation. A RDA tem como principais objetivos o desenvolvimento e adoção de infraestruturas que possibilitem a partilha de dados, desenvolvendo soluções capazes de enfrentar os desafios inerentes à partilha, interpretação e reutilização dos dados. O JISC<sup>2</sup> é outro exemplo, uma organização sem fins lucrativos que desenvolve soluções para a educação e investigação no Reino Unido. Um exemplo dessas soluções é o curso MANTRA<sup>3</sup>, financiado pelo JISC, desenvolvido pela Universidade de Edimburgo, um curso online gratuito que apoia todos os que necessitem de gerir dados de investigação.

Também a Comissão Europeia tem encorajado o acesso aberto e a reutilização dos dados de investigação produzidos pelos projetos Horizonte 2020 através do Open Research Data Pilot (ORD Pilot), de acordo com os princípios dos dados FAIR - todos os dados de investigação devem ser *Findable, Accessible, Interoperable and Reusable*. (European Research Council 2017). O problema reside nas pequenas comunidades, referidas frequentemente por “*long tail*” ou cauda longa, que por norma, não têm soluções satisfatórias que permitam tratar dos dados de modo que estes possam ser armazenados, difundidos, facilmente acedidos e compreendidos por outros investigadores.

O valor dos dados de investigação está diretamente relacionado com a estrutura e organização dos mesmos pelo que é fundamental a realização da gestão dos dados de investigação a par do processo de investigação, de forma a garantir o retorno do investimento. De acordo com Assante et al. (2016) o custo é um dos principais fatores que impede que a publicação dos dados seja uma norma para a ciência, incluindo também o esforço necessário para preparar os conjuntos de dados, de forma a que outros possam fazer uso dos mesmos. Os esforços relacionam-se com a formatação e documentação dos dados por sua vez, os custos estão ligados ao depósito dos dados em repositórios confiáveis que tornem os dados acessíveis.

---

<sup>1</sup> RDA- <https://rd-alliance.org/> acedido em: 3 de janeiro de 2018

<sup>2</sup> JISC- <https://www.jisc.ac.uk/> acedido em: 3 de janeiro de 2018

<sup>3</sup> MANTRA- <https://mantra.edina.ac.uk/> acedido em: 3 de janeiro de 2018

A curadoria de dados não é uma prática comum das instituições, assim na maioria dos casos os investigadores armazenam os dados sem que estes estejam documentados (Castro et al. 2015) comprometendo a partilha e respetiva reutilização por outros investigadores “aumentando o risco de perda de dados” (Smit, Van der Hoeven, e Giaretta 2011).

São muitos os benefícios da partilha de dados de investigação, segundo a OECD (2007) o acesso aos dados de investigação aumenta os retornos do investimento público; reforça investigação científica aberta; incentiva a diversidade de estudos e opiniões; promove novas áreas de trabalho e permite a exploração de novos tópicos. A Comissão Europeia, (2017) identifica como benefícios da partilha de dados de investigação a melhoria da qualidade dos resultados, uma vez que os investigadores podem aceder a investigações anteriores; o incentivo da colaboração evitando a duplicação de esforços, acelerando assim a inovação e o desenvolvimento científico; e o aumento da transparência do processo científico. A UK Data Archive (2011) acrescenta ainda como benefícios o incentivo à melhoria e à validação de métodos de pesquisa; o aumento do impacto e da visibilidade da pesquisa, bem como a promoção do autor, que recolheu/ criou os dados, através do aumento das citações. Assim, pode concluir-se que a partilha de dados é um aspeto positivo para a comunidade científica, no entanto de acordo com Castro (2013) os benefícios são menos óbvios para quem produz e torna os dados disponíveis, possivelmente pelo facto de a investigação se tratar de um processo que consome vários recursos. À medida que a difusão de conteúdos seja facilitada e os benefícios se tornem mais óbvios para o investigador, é possível que os autores se tornem mais aptos a partilhar os dados dos seus estudos e dessa forma potencializar a sua utilidade para a sociedade.

O RCAAP (2010) identifica barreiras à partilha dos dados de investigação como a cultura da reutilização de dados resultantes de outras investigações não se encontrar, ainda, enraizada na comunidade de investigadores, pelo que os benefícios explícitos da disponibilização e da partilha de dados científicos são ainda muito escassos ou inexistentes. A recolha e tratamento de dados exige muito esforço e processa-se à custa de conhecimentos especializados, assumindo, com frequência, o estatuto de investimento de longa duração. O receio que os dados produzidos sejam explorados e reutilizados de forma incorreta faz com que alguns investigadores rejeitem a reutilização de dados recolhidos no âmbito de outros processos de investigação, atendendo às diferenças que existem na conceção de processos experimentais e na recolha de dados.

São poucos os investigadores com a experiência e o domínio de competências técnicas para a organização e o processamento dos dados, embora a vertente mais ignorada seja, ainda, a da criação de metadados capazes de garantir o acesso aos dados durante o seu ciclo de vida, bem como de assegurar a sua preservação a longo termo. Acresce ainda o receio das despesas resultantes do investimento em gestão de dados, bem como a falta de tempo para desenvolver a investigação quando sobrecarregados com outras tarefas. É ainda apontado como um grande

*desafio a identificação dos dados que podem ser partilhados, por quem, com quem, quais as condições, porquê e quais são os efeitos da partilha (Borgman 2012).*

A dificuldade que os investigadores sentem para documentar os seus conjuntos de dados é uma das razões apontadas para a relutância na partilha de dados de investigação. Considerando a quantidade de tarefas com que se teriam de preocupar para que os dados pudessem ser bem descritos, com metadados associados capazes de possibilitar a partilha e recuperação. Além da necessidade de encontrar ferramentas, modelos que lhes permitissem garantir o tratamento, armazenamento, disseminação e preservação a longo prazo dos dados de investigação. Assim, para garantir *“que os dados possam ser partilhados e reutilizados é necessário que existam organizações e indivíduos responsáveis pela sua curadoria. Os investigadores, por si só, não serão as pessoas adequadas para assegurar a preservação e o acesso continuado aos dados que recolhem e produzem”*(RCAAP 2010). De acordo com a NSF (2005) os gestores de dados e as organizações são os responsáveis pela operação de manutenção do conjunto de dados, frequentemente o processo de depositar dados numa coleção é uma responsabilidade partilhada entre os investigadores e os gestores de dados, embora por vezes a partilha de responsabilidades varie.

São vários os *stakeholders* envolvidos na gestão de dados de investigação, desde o investigador, à instituição/organização responsável pelo projeto de investigação, ao financiador, ao gestor de dados, ao utilizador, entre outros. É recorrente que a partilha de responsabilidades varie entre os envolvidos. Dessa forma, de acordo com (Lyon 2007) é da responsabilidade dos investigadores, bem como da instituição e do gestor de dados, trabalhar os dados resultantes das investigações de forma a garantir que podem ser usados por outros. No caso das organizações, ou projetos em que a mão de obra especializada é reduzida, ou os investimentos são menores, cabe aos investigadores suportar as tarefas relacionadas com a descrição dos dados de investigação. Por outro lado, se existirem gestores de dados, os investigadores devem ser envolvidos no processo de descrição uma vez que têm o conhecimento sobre domínio, compreendendo quais os conceitos fundamentais que necessitam ser descritos, para que os dados possam ser interpretados por outros. *“Os conjuntos de dados tendem a ser únicos sem múltiplas cópias, não se descrevem por si só, podem exigir muita interpretação, competências específicas e conhecimento de domínio”*(Lyon 2007). Por esta razão se apresenta preponderante a interação entre os envolvidos na RDM, a responsabilidade é partilhada entre as comunidades científicas, os financiadores, as instituições e os gestores de dados que devem trabalhar em conjunto de modo a desenvolverem soluções, para que seja possível interpretar e reutilizar os dados a longo prazo.

Outro aspeto importante a considerar é referido por (Wilson 2007), que afirma que os profissionais responsáveis pelo desenvolvimento de metadados são proficientes. Contudo são poucos para conseguirem dar resposta em tempo útil ao tratamento dos conjuntos de dados,

por essa razão por vezes abdica-se da qualidade dos metadados fazendo uso das descrições efetuadas pelos criadores dos conjuntos de dados, ou utilizando processos automáticos para o desenvolvimento de metadados. Chao (2014) apresenta como uma solução alternativa viável a análise de documentos, produzidos por investigadores. Da análise desses documentos é possível extrair informação para o desenvolvimento de metadados. Este método garante ao curador de dados uma maior autonomia, garantindo que possa ser desenvolvido um trabalho preliminar de análise do domínio e desenvolvimento do modelo de metadados antes da interação com o investigador.

A qualidade dos metadados pode ser influenciada pela falta de sensibilidade dos investigadores para a importância da gestão dos dados de investigação. Considerando que os dados são seus e a consulta dos mesmos apenas será efetuada por eles no decorrer da investigação, e por essa razão descrevem os conjuntos sem a preocupação de que outros compreendam o que está representado. Um exemplo simples é a utilização de títulos como “exrefluz” (experiência sobre a refração da luz), este título não será compreendido por terceiros, dessa forma não será uma boa forma de nomear o documento.

*“Os dados existem apenas aos olhos de quem os vê, por vezes os investigadores armazenam coleções de materiais sem compreender o quão valioso eles podem ser como dados”* (Borgman 2012). Ou seja, os dados só são valiosos se alguém identificar alguma potencialidade nos mesmos, essa interpretação irá depender do conhecimento de cada pessoa, o que aos olhos de um investigador não tem qualquer relevância, para outro investigador pode ser a chave para uma nova investigação.

## **2.1 Dados de investigação**

Os dados de investigação são recursos valiosos, produzidos ou utilizados no contexto de investigação científica. É com base na análise dos dados que os investigadores apoiam as suas decisões e retiram conclusões que incentivam a inovação e permitem o avanço do conhecimento científico. Porém o valor dos dados pode ser influenciado pela estrutura e organização dos mesmos, sendo potencializado quando disponibilizados no meio digital, permitindo que outros possam ter acesso, incentivando a geração de novo conhecimento científico e fomentando o trabalho colaborativo entre domínios diversos.

A OECD (2007) no contexto do “*Principles and Guidelines*” define os dados de investigação como registos factuais que podem ser (números, registos textuais, imagens e sons) utilizados como fontes primárias para as investigações científicas, sendo habitualmente aceites na comunidade científica como recursos necessários para a validação de resultados das

investigações. Também na definição do dicionário Casrai<sup>4</sup> os dados de investigação são fontes primárias que suportam as investigações e são usados como evidência no processo de investigação sendo imprescindíveis para validar os resultados. Nesta definição é referido ainda que qualquer conteúdo em formato digital ou analógico tem potencial para ser dado de investigação. Por sua vez, a Comissão Europeia (2017) refere-se aos dados de investigação como informações, em particular factos ou números, recolhidos com o objetivo de serem analisados e considerados como base para o desenvolvimento da fundamentação, discussão ou cálculo das investigações científicas. Acrescenta ainda como exemplos de dados em contexto de investigação as estatísticas, resultados de experiências, os registos de entrevistas e as imagens.

Os dados científicos podem assumir diferentes classificações de acordo com a origem; nível de tratamento que sofreram; área científica que os produziu ou utilizou e ainda uma classificação desenvolvida pela CODATA em nome da UNESCO (Willis Craig, Greenberg, White 2012) que agrupa os dados em quinze categorias.

A classificação dos dados científicos de acordo com Willis Craig, Greenberg e White (2012), é extremamente útil na medida em que possibilita a compreensão das semelhanças e das diferenças entre conjuntos de dados, facilitando ainda o entendimento do uso potencial dos dados de investigação ao longo do tempo. Note-se que segundo a NSF (2005) a natureza dos dados de uma coleção pode ser diversa, podendo incluir simultaneamente números, imagens, vídeos ou áudio, software e informações sobre as versões, algoritmos, equações, animações ou modelos.

A NSF (2005) atribuiu três classificações distintas de acordo com a origem dos dados de investigação, assim podem ser:

**observacionais**- registos históricos que não podem voltar a ser recolhidos e que por essa razão deverão ser arquivados indefinidamente (observações diretas da temperatura do oceano numa data específica, medidas de precipitação, registos fotográficos de uma atividade vulcânica);

**computacionais**- resultado da execução de modelos computacionais ou de simulações podem ser reproduzidos se tiverem associada informação completa sobre o hardware, software e os *inputs*, uma vez ser possível a reprodução do modelo ou simulação, que deu origem aos dados. A sua preservação num repositório pode não ser necessária, embora seja preponderante o arquivo do modelo e do conjunto de metadados associados.

**experimentais**- normalmente são possíveis de reproduzir, como por exemplo as taxas de reação química ou desempenho de um motor. Por essa razão não será necessário armazenar indefinidamente, contudo a opção do armazenamento poderá variar de acordo com o custo da

---

<sup>4</sup> Casrai. (2017) Research data. [28 de dezembro de 2017] Disponível em: [http://dictionary.casrai.org/Research\\_data](http://dictionary.casrai.org/Research_data)

reprodução da experiência. Por vezes não é possível reproduzir todas as condições, particularmente quando as condições ou as variáveis são desconhecidas, ou quando os custos de reprodução são avultados, nestes casos será necessário garantir a preservação a longo prazo.

De acordo com Willis Craig, Greenberg e White (2012) os dados experimentais resultantes de sistemas bem definidos podem ser verificados e reproduzidos independentemente do tempo ou local, desde que os procedimentos e as variáveis relevantes estejam bem documentados. Este tipo de dados resulta normalmente de disciplinas como a física, termodinâmica, química, biologia molecular ou estrutural. Por outro lado, os dados observacionais não podem ser reproduzidos, recolhidos ou verificados, uma vez que são normalmente dependentes do tempo e lugar, são frequentemente produzidos por disciplinas como a biologia, ecologia e ciências sociais.

Os modelos computacionais ou simulacionais também podem ser recriados e verificados desde que tenham metadados associados. *“The cost of model validation is usually quite significant, especially when extremely high model confidence is required”* (Sargent 2011). Os modelos simulacionais devem ser validados para poderem ser utilizados, o facto de o modelo simulacional ser válido não significa que este seja capaz de dar resposta a todas as questões do domínio em causa, é válido para dar resposta a um conjunto de questões definidas. Assim de acordo com Sargent a validação e verificação do modelo deve ser parte do processo do modelo simulacional. Dado o custo de validação de um modelo ser muito elevado, afirma-se a necessidade de garantir a recuperação acessibilidade e preservação a longo prazo, reforçando a importância de gerir os dados de investigação. Algumas das técnicas de validação usam a comparação com outros modelos anteriormente validados, pelo que é notória a necessidade da partilha de dados de investigação, para que seja possível realizar este tipo de comparações.

Segundo Westbrook e Gattidge (1991) referido por (Willis Craig, Greenberg, White 2012) o glossário CODATA defende que os dados experimentais podem ser considerados como dados não validados ou dados validados quando sujeitos a processos de validação, ou gerados de acordo com métodos padrão.

De acordo com CODATA citada por Willis Craig, Greenberg e White (2012) e Lauriault et al. referido por Castro (2013) os dados de investigação, de acordo com o nível de processamento, podem ser classificados como dados brutos (*raw data*) dados que não foram processados, ou dados processados, relativos a dados em bruto que já foram sujeitos a um processo de manipulação.

JISC<sup>5</sup> sugere que a origem dos dados de investigação que contemplam a experimentação, simulação ou observação diferem significativamente entre disciplinas e ainda entre disciplinas do mesmo domínio.

A OECD (2015) refere que vários investigadores e utilizadores de dados consideram relevante classificar as unidades de investigação e desenvolvimento, possibilitando assim a distinção entre os recursos produzidos. A classificação que se propõe é a de FORD referida pela OECD (2015) que divide as unidades de investigação em seis grandes categorias que são: Ciências Naturais; Engenharia e Tecnologia; Medicina e Ciências da Saúde; Agricultura e Ciências Veterinária; Ciências Sociais; Humanidades e Artes, cada uma delas apresenta subdivisões tal como se pode verificar na Figura 1.

Broad classification	Second-level classification
1. Natural sciences	1.1 Mathematics 1.2 Computer and information sciences 1.3 Physical sciences 1.4 Chemical sciences 1.5 Earth and related environmental sciences 1.6 Biological sciences 1.7 Other natural sciences
2. Engineering and technology	2.1 Civil engineering 2.2 Electrical engineering, electronic engineering, information engineering 2.3 Mechanical engineering 2.4 Chemical engineering 2.5 Materials engineering 2.6 Medical engineering 2.7 Environmental engineering 2.8 Environmental biotechnology 2.9 Industrial biotechnology 2.10 Nano-technology 2.11 Other engineering and technologies
3. Medical and health sciences	3.1 Basic medicine 3.2 Clinical medicine 3.3 Health sciences 3.4 Medical biotechnology 3.5 Other medical science
4. Agricultural and veterinary sciences	4.1 Agriculture, forestry, and fisheries 4.2 Animal and dairy science 4.3 Veterinary science 4.4 Agricultural biotechnology 4.5 Other agricultural sciences
5. Social sciences	5.1 Psychology and cognitive sciences 5.2 Economics and business 5.3 Education 5.4 Sociology 5.5 Law 5.6 Political science 5.7 Social and economic geography 5.8 Media and communications 5.9 Other social sciences
6. Humanities and the arts	6.1 History and archaeology 6.2 Languages and literature 6.3 Philosophy, ethics and religion 6.4 Arts (arts, history of arts, performing arts, music)

FIGURA 1-CLASSIFICAÇÃO DAS ÁREAS DE INVESTIGAÇÃO SEGUNDO FORD

FONTE: (OECD 2015)

<sup>5</sup> JISC. (2018) Joint Information Systems Committee. [4 de janeiro de 2018] Disponível em: <https://www.jisc.ac.uk/>



A produção de dados de investigação e consequente novo conhecimento em determinadas áreas científicas, dá origem em muitos casos ao desenvolvimento de novos esforços científicos em áreas diferentes, uma vez que a partilha de conhecimento científico é a base para a inovação e o desenvolvimento tecnológico. Tomando como exemplo o uso de elementos tecnológicos como forma de prevenção e tratamento de patologias na área da Medicina.

Tendo por base um estudo desenvolvido pela CODATA em nome da UNESCO Kotani (1975) citado por Willis Craig, Greenberg e White 2012 desenvolveu uma classificação de dados de investigação. Esta classificação agrupa os dados em quinze categorias: “dados que podem ser mensurados repetidamente” ou “dados que podem ser medidos apenas uma vez”, "independente do local" ou "dependente da localização", a Figura 2 lista as quinze categorias com a respetiva descrição e exemplos associados.

*Classification of data from Kotani (1975)*

Categories of data	Description	Examples
Time-independent	Data that can be measured repeatedly	Most data in chemistry and physics, geological structures
Time-dependent	Data that can be measured only once	Volcanic eruptions, rare specimens, fossils
Location-independent	Data independent of location of objects measured	Most data in chemistry and physics, minerals, most data in biological sciences
Location-dependent	Data dependent on location of objects measured	Rocks, fossils, astronomical data, meteorological data, rare specimens, fossils
Primary observational or experimental data	Data obtained by experiment or observation	Optical spectra, crystallographic $F$ values, seismographic records, weather charts
Derived data	Data derived by combining several primary data with the aid of a theoretical model	Fundamental constants, crystal structures, temperature distribution
Theoretical data	Data derived by theoretical calculations	Predicted solar eclipses
Determinable data	Data on a quantity which can be assumed to take a definite value under a given condition	Most macroscopic data
Stochastic data	Data on a quantity that takes fluctuating values from one sample or measurement to another	Polymer data, structure-sensitive properties, soil composition, solar flares, most metrology
Quantitative data	Measures of scientific quantities in terms of well-defined units	Most data in chemistry and physics, seismic data, meteorological data
Semiquantitative data	Measures of scientific quantities using arbitrary scales	Mohs hardness scale, wind force scale
Qualitative data	Any scientific definitive statement concerning scientific objects	Chemical structure formulas, properties of nuclides, rock classification, amino-acid sequences
Data as numerical values	Data presented as isolated numerical values	Meteorological data
Data as models or graphs	Data presented in graphical form or as models	Phase diagrams, molecular models, geologic maps, genetic pathways
Symbolic data	Data presented using arbitrary symbols (non-numeric)	Lithology in borehole data

FIGURA 2-CLASSIFICAÇÕES DE DADOS KOTANI (1975)

FONTE: OECD (2015)

Através da análise da figura anterior verifica-se que grande parte dos dados produzidos na física e na química são independentes do tempo, ou seja, podem ser medidos repetidamente em momentos diferentes. Induz-se que são essencialmente dados do tipo experimental ou simulacional. Por outro lado, os dados resultantes de estudos sobre erupções vulcânicas, fósseis são dependentes do tempo, por essa dependência apenas podem ser medidos uma vez. Um exemplo são os dados registados da atividade do vulcão Etna no dia 8 de março de 2017, os dados medidos posteriormente nunca serão os mesmos registados nessa data, assim este tipo de investigações gera dados essencialmente do tipo observacional.

Grande parte dos dados resultantes de investigações nas ciências biológicas, na física, química, são independentes do local, logo serão essencialmente dados do tipo experimental ou simulacional. Por sua vez, os dados registados sobre astronomia, meteorologia são dependentes da localização, por exemplo a precipitação registada num determinado local não será a mesma registada num outro local distinto, infere-se que os dados serão essencialmente do tipo observacional.

As várias classificações de dados de investigação e as suas origens contextuais sugerem que os metadados, usados para documentar o conteúdo, a estrutura e o contexto terão características diferentes que se relacionam com cada tipologia de dados, este aspeto influencia os requisitos a ter em consideração aquando do desenvolvimento de metadados para a reutilização de dados de investigação.

Este trabalho foca os dados experimentais, em que é possível a sua reprodução caso exista informação sobre variáveis, metodologia de recolha e configuração das experiências. O suporte à descrição de dados é escasso, o que inibe a produção de metadados, inviabiliza a sua reprodução e a perda dos dados é, na maioria dos casos, inevitável.

A maximização do potencial dos dados de investigação só é possível quando os investigadores partilham os seus dados em plataformas para o efeito, com um conjunto de metadados adequados.

## **2.2 Plataformas de Gestão de dados de investigação**

São diversas as plataformas para comunidades bem definidas, como são o caso a Europe PubMed Central para as ciências da vida, o British Atmospheric Data Center para as ciências ambientais e SBGrid para a biologia estrutural. Contudo, na cauda longa da ciência, os investimentos e as ferramentas são modestas e na maioria dos casos não se ajustam às necessidades.

Segundo Assante et al. (2016) a publicação de dados de investigação pretende divulgar os dados para possibilitar aos profissionais a (re)utilização de acordo com a dinâmica da “ciência aberta”. Amorim et al. (2016) acreditam que a gestão de dados de investigação começa a tornar-se uma preocupação recorrente dos investigadores e nesse sentido, as instituições necessitam fornecer plataformas que apoiem a organização e partilha dos dados. Algumas instituições optaram por repositórios institucionais outras optam por ambientes mais ricos de descrição de dados. Os repositórios institucionais apresentam-se assim como uma solução que garante a recuperação, acesso e preservação a longo prazo de grandes quantidades de dados, sem custos.

Heidorn (2008) referido por Assante et al. (2016) defende que a demanda por repositórios na cauda longa da ciência está a emergir, no contexto destes domínios científicos,

em que a atividade é desenvolvida por um grande número de laboratórios relativamente pequenos e por investigadores individuais que produzem coletivamente a maioria dos resultados científicos.

Apesar da procura por soluções ser cada vez maior, a dificuldade prende-se com a falta de capacidade, por parte dos repositórios centrais, em apoiar o depósito de dados de investigação provenientes de domínios diversos. A tentativa de utilizar modelos de dados genéricos para a descrição de conjuntos de dados heterogéneos, provenientes de vários domínios, causam lacunas na gestão, recuperação e consequentemente inviabiliza a reutilização dos dados. Tal como refere Assante et al. (2016) os repositórios usam um pequeno conjunto de modelos de metadados para lidar com domínios diversos.

Assim, considera-se que as soluções dos repositórios de dados estão aquém do desejável, reforçando-se a importância do desenvolvimento de modelos de metadados adequados às necessidades de domínios específicos. Assante et al. (2016) acrescenta ainda que muitos dos desafios dos repositórios têm a ver com a descrição e a preservação a longo prazo, assim devem ser clarificadas pelo menos duas diferenças fundamentais entre as publicações e os conjuntos de dados: estes últimos são muitas vezes puramente numéricos, o que torna muito difícil a atribuição de qualquer tipo de metadados, simplesmente através da análise dos números. Além disso, os conjuntos de dados exigem que as descrições detalhadas e específicas do domínio sejam interpretadas corretamente. Outro aspeto importante é a questão dos requisitos de metadados que também podem variar muito de domínio para domínio, exigindo que os modelos de dados do repositório sejam flexíveis o suficiente para representar os dados de investigação.

O facto de ainda não existirem soluções que se adequem totalmente à gestão e respetiva disponibilização e (re)utilização dos dados, não deve ser inibidor da partilha de dados. O projeto Re3Data<sup>6</sup> poderá ser uma “ferramenta” extremamente útil, uma vez que se trata do maior registo de repositórios de dados disponibilizado na Web, na sua lista constam mais de 1500 registos de repositórios de dados de investigação. Assim este diretório pode servir como um ponto de partida para a identificação de repositórios adequados à área de interesse. De acordo com European Research Council (2017) se não existir um repositório específico que se adequa aos dados gerados por uma determinada disciplina, os investigadores devem disponibilizar as suas publicações em repositórios institucionais ou em sistemas centralizados, como é o caso do Zenodo.

Este tipo de repositórios foi desenvolvido com o objetivo de auxiliar a publicação de conjuntos de dados produzidos por domínios científicos na “cauda longa”.

Nas subseções seguintes apresentam-se alguns repositórios e o diretório Re3Data.

---

<sup>6</sup> Re3Data- <https://www.re3data.org/> acedido em: 11 de janeiro de 2018

### **2.2.1 Zenodo**

O Zenodo<sup>7</sup> é um repositório, lançado em maio de 2013, desenvolvido pelo CERN através da iniciativa European Union OpenAIREplus. Tem como objetivo garantir que pequenos projetos científicos possam partilhar os seus objetos digitais resultantes do processo de investigação, de forma gratuita. Aceita conteúdos de todos os domínios e nações, não impondo qualquer requisito de formato, tamanho, restrições de acesso ou licença. Garante a possibilidade de armazenar conteúdo não publicado assegurando que os dados são capturados e armazenados de forma segura enquanto a pesquisa está em desenvolvimento.

O Zenodo apoia os investigadores a obter crédito tornando possível a citação dos resultados de investigação através do OpenAIRE. Outro aspeto interessante é o facto de o código do repositório ser aberto, construído sobre a base da biblioteca digital Invenio.

O acesso aberto ao código pode ser um dos critérios para a seleção de um sistema em relação a outro, uma vez que se trata da garantia de que o repositório não fica dependente do fornecedor, normalmente associado a software e outros serviços que são pagos posteriormente. Evita-se também problemas caso o fornecedor se torne incapaz de suportar o repositório e ainda possibilita efetuar modificações, ajustando-se às necessidades da organização que adquire o sistema.

No que diz respeito ao tipo de metadados usado para a descrição dos objetos a plataforma disponibiliza alguns dos metadados apresentados: Título; Autor; Data de Publicação; DOI; Comunidade; Tipo de Licença; Assunto(s) tal como se pode verificar na Figura 3 e Figura 4.

---

<sup>7</sup> Zenodo- <https://zenodo.org/> acedido em: 12 de janeiro de 2018

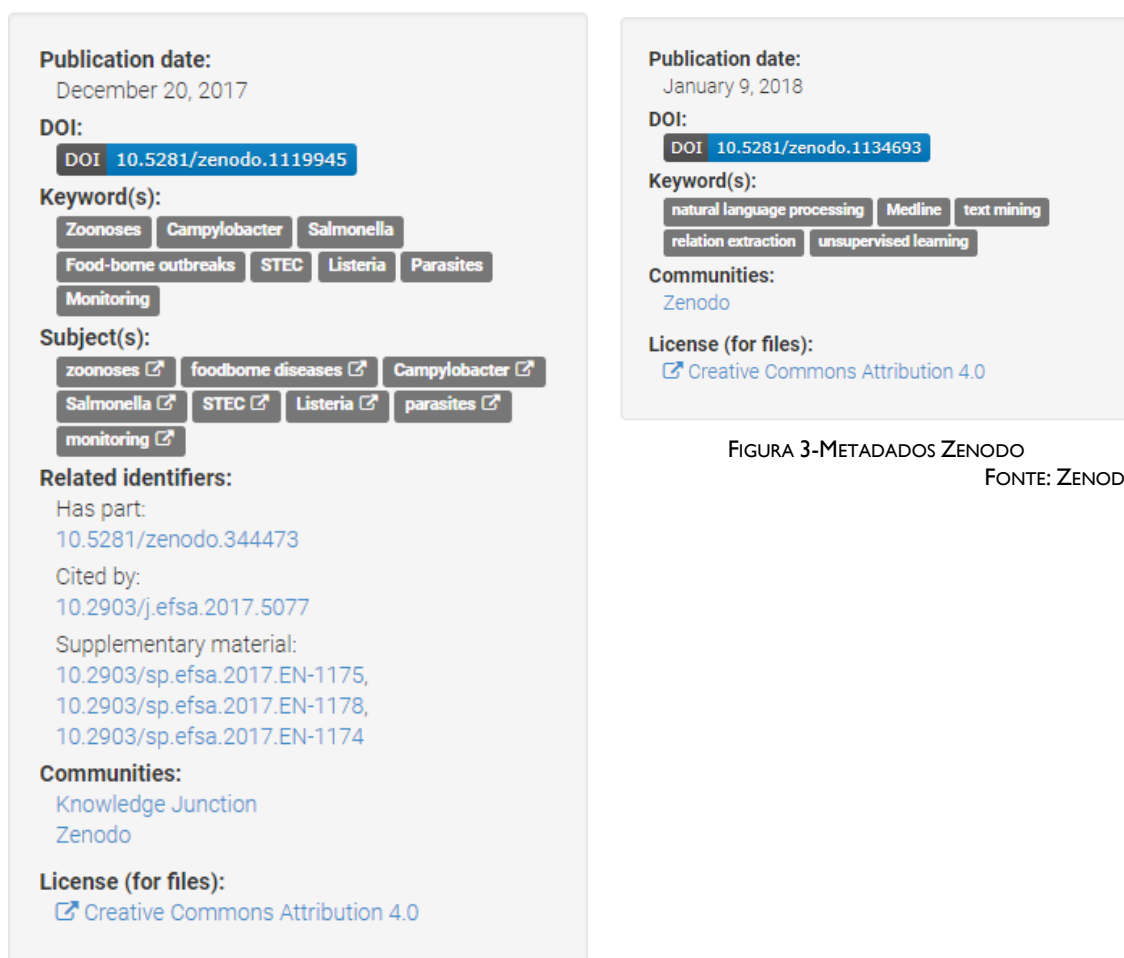


FIGURA 3-METADADOS ZENODO  
FONTE: ZENODO

FIGURA 4-METADADOS ZENODO  
FONTE: ZENODO

### 2.2.2 Figshare

O Figshare<sup>8</sup> foi desenvolvido por Mark Hahnel, que defendia o conceito de dados abertos. Este repositório é financiado pela Digital Science, divisão da tecnologia global da Macmillan Science & Education. Tem como principal objetivo possibilitar o armazenamento, gestão e divulgação de qualquer tipo de resultado de investigação. Trata-se de um repositório pago, cujos valores variam de acordo com a quantidade de artigos; tamanho dos dados; e a complexidade de implementação e integração com outros sistemas. Garante que os autores partilhem facilmente os seus resultados sem se terem de preocupar com o tamanho ou o formato dos objetos; é possível a citação de todos os dados, sendo que estes têm um DOI (Digital Object Identifier) associado.

<sup>8</sup> Figshare- <http://figshare.com> acedido em: 12 de janeiro de 2018

Relativamente aos metadados é possível verificar que usam para a descrição dos objetos: Título; Autor; Descrição; Data; Palavras chave; Licença; Assunto(s) tal como apresentado na Figura 5.

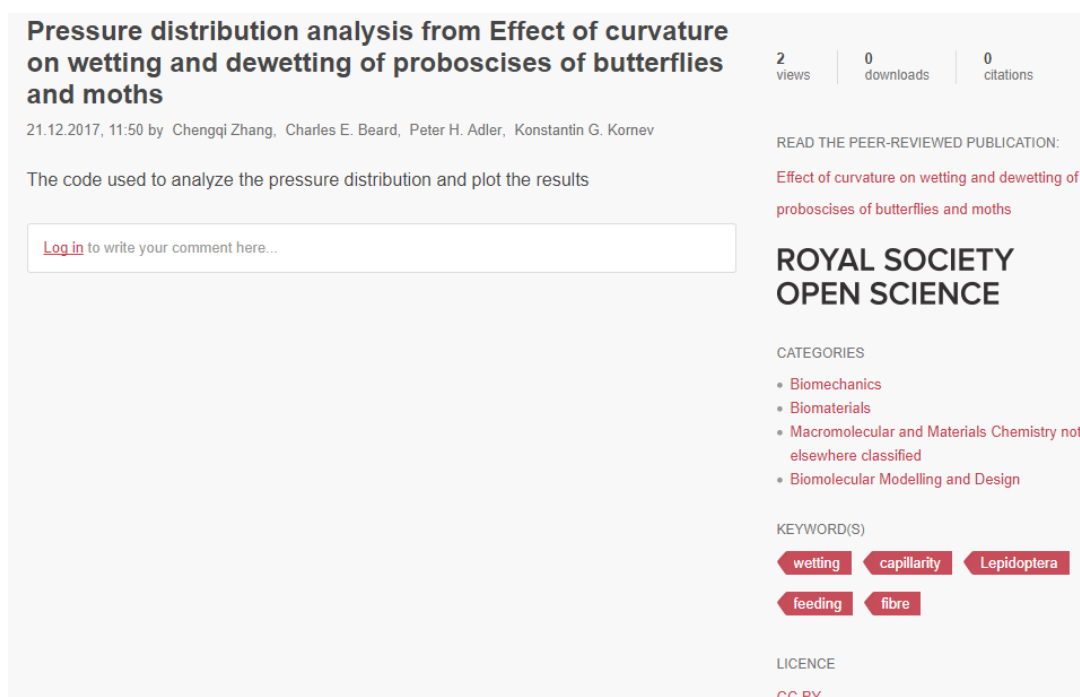


FIGURA 5-METADADOS FIGSHARE

FONTE: FIGSHARE

### 2.2.3 Dryad

O Dryad <sup>9</sup> é um repositório sem fins lucrativos que resulta da iniciativa conjunta entre um grupo de revistas e comunidades científicas (biologia evolutiva e ecologia), tinham como objetivo desenvolver uma solução sustentável para o tratamento de dados promovendo a ciência aberta e consequentemente a partilha do conhecimento. Estabelece parceria com o DataONE de forma a garantir o acesso aos conteúdos indefinidamente. Este repositório possibilita a partilha de qualquer tipo de formato de dados; permite o armazenamento de publicações e dados mantendo-os privados durante o período necessário para se efetuarem as revisões e estabelecerem os períodos de embargo.

O Dryad apresenta algumas preocupações com a interpretação e reutilização, por essa razão encoraja os autores a fornecer documentação adicional que permita descrever de forma eficiente o conjunto de dados; atribui aos dados um DOI garantindo a possibilidade de citação dos dados e dessa forma os autores obterem crédito. Os conteúdos disponibilizados são gratuitos, podendo ser reutilizados sem qualquer tipo de barreiras.

<sup>9</sup> Dryad- <http://datadryad.org/> acedido em: 12 de janeiro de 2018

Para descrever os seus recursos usa frequentemente os seguintes metadados: Título; Autor; Data de Publicação; DOI; Descrição, “Downloaded” (número de vezes que efetuaram download) ver Figura 6.

The screenshot displays the Dryad interface for a dataset. At the top, the title 'Data from: Rapid responses of soil microorganisms improve plant fitness in novel environments' is shown, followed by the authors 'Lau JA, Lennon JT', the publication date 'Date Published: February 7, 2013', and the DOI 'https://doi.org/10.5061/dryad.qc537'. Below this, a section titled 'Files in this package' includes a disclaimer about the 'as is' nature of the data and a Creative Commons license. A table lists the dataset details:

Title	Lau and Lennon PNAS soil N and microbe data
Downloaded	82 times
Description	This csv file contains the microbial community composition and abundance data.
Download	<a href="#">Lau and Lennon PNAS microbe data.csv (3.328 Kb)</a>
Download	<a href="#">README.txt (6.978 Kb)</a>
Details	<a href="#">View File Details</a>

FIGURA 6-METADADOS DRYAD

FONTE: DRYAD

#### 2.2.4 CKAN

CKAN<sup>10</sup> (Comprehensive Knowledge Archive Network) foi desenvolvido pela Open Knowledge Foundation em 2014 é um sistema de código aberto e sem fins lucrativos. Foi desenvolvido com o objetivo de tornar conteúdo público e aberto em todos os países, possibilitando a partilha de forma gratuita. O repositório aceita conteúdo de vários formatos; aos recursos partilhados podem ser agregados vários tipos de objetos digitais, ou ainda links para páginas externas (jornal onde o documento ou o conjunto de dados possa ter sido partilhado). A cada recurso é associado um conjunto de metadados, sendo que a configuração padrão engloba: Autor, Título, Descrição, Licença, Data de criação. Alguns destes metadados podem ser identificados na Figura 7.

<sup>10</sup> CKAN- <https://ckan.org/> acedido em: 16 de janeiro de 2018

Dataset
Groups
Activity Stream

## TEST REUS

Prova de georeferenciado de dades

### Data and Resources

csv

REUS\_activitats

Test georeferenciado llicencies activitats Anatomy: working shopping

Explore

function indicators

### Additional Info

Field	Value
Author	Antonio Ripoll
Last Updated	10 de Janeiro de 2018, 18:01 (UTC+00:00)
Created	3 de Dezembro de 2017, 10:00 (UTC+00:00)

FIGURA 7-METADADOS CKAN

FONTE: CKAN

## 2.2.5 B2SHARE

O B2SHARE<sup>11</sup> foi lançado pela iniciativa EUDAT com o objetivo de fornecer aos investigadores uma solução fiável, para o armazenamento e partilha de conjuntos de dados de investigação, de pequena escala. O repositório é gratuito e permite a atribuição de DOI aos conjuntos de dados, garantindo o acesso duradouro e a possibilidade de referência. O B2SHARE garante a integridade dos dados durante a ingestão dos mesmos, por outro lado é da responsabilidade do autor definir a política de acesso aos recursos que ingere no sistema; o autor tem à sua disposição um conjunto de metadados que pode usar para a descrição dos conteúdos que partilha. Alguns dos metadados usados são: Autor, Data de Publicação; Descrição; Palavras-chave; DOI tal como se pode ver na Figura 8.

### Costs of FAIR Compliance and not being FAIR compliant

by Peter Wittenburg:

Jan 7, 2018

**Description:** This is a short note about costs for creating FAIR compliant data accepting that it is hardly possible to estimate costs for not being FAIR compliant. This note grew over time based on a request from people making a survey about costs and comments from colleagues. It is mainly based on own experience.

**Keywords:** FAIR, costs, data, inefficiencies, infrastructure;

**DOI:** 10.23728/b2share.e184bd1f12d45269de80c3f3e443eb7 Copy

**PID:** 11304/ca773b0c-2dfe-4071-bb8a-2fcb76b6a867 Copy

Files

Name	Size
> funding-fair.pdf	968,73KB

Basic metadata

Open Access

True ✓

Contact Email

Peter.wittenburg@mpcdf.mpg.de

FIGURA 8-METADADOS B2SHARE

FONTE: B2SHARE

<sup>11</sup> B2SHARE- <https://b2share.eudat.eu/> acedido em: 16 de janeiro de 2018



### 2.2.6 Re3Data

O Re3Data<sup>12</sup> é atualmente o maior registo de repositórios de dados disponibilizado na Web. Este diretório é um serviço do DataCite<sup>13</sup> administrado por uma colaboração internacional entre Purdue University dos Estados Unidos e pelo Karlsruhe Institute of Technology da Alemanha.

Atualmente conta com mais de 1500 registos de repositórios de dados de investigação de vários domínios desde ciências naturais; humanidades; engenharia entre muitos outros. O Re3Data permite que os investigadores, as agências de financiamento, estudantes, ou a comunidade em geral encontrem os repositórios de dados mais apropriados para depositar, ou aceder aos dados resultantes de investigações científicas.

Relativamente à pesquisa pode ser efetuada de várias formas, uma vez que são disponibilizados vinte e sete filtros: por assuntos; tipo de conteúdo; por país; políticas; locais, entre outros.

Por norma, os repositórios lidam com um conjunto reduzido de metadados para efetuarem a descrição dos objetos digitais. Usando essencialmente o Título, o Autor; a Data de publicação, uma breve descrição e em alguns casos apresentam também o DOI, e o Assunto.

Esta breve análise leva à confirmação do que foi referido no início deste capítulo. Os repositórios centrais, ainda não têm soluções para apoiar o depósito de dados de investigação de domínios diversos. Os modelos de metadados genéricos não são suficientes para a descrição de conjuntos de dados com contextos de produção distintos.

A solução pode passar pela utilização de ferramentas que possibilitem a organização dos dados desde a sua produção, garantindo ainda um meio flexível para a combinação de metadados. O Dendro<sup>14</sup> é um bom exemplo disso, uma plataforma colaborativa de gestão de dados, desenvolvida pelo Laboratório de I&D de Sistemas de Informação da Faculdade de Engenharia da Universidade do Porto. Esta plataforma permite adicionar metadados ajustados às necessidades específicas de domínios diversos.

## 2.3 Metadados para a descrição de dados de investigação

A definição literal para metadados, com base na etimologia da palavra é “dados sobre dados” (Riley 2017). Os metadados são dados desenvolvidos com o objetivo de garantir a autenticidade de outros dados; atribuindo-lhes informação que permite aumentar a

---

<sup>12</sup> Re3Data- <https://www.re3data.org/> acedido em: 17 de janeiro de 2018

<sup>13</sup> DataCite- <https://www.datacite.org/> acedido em: 17 de janeiro de 2018

<sup>14</sup> Dendro- <http://dendro-prd.fe.up.pt:3007/> acedido em: 17 de janeiro de 2018

inteligibilidade garantindo assim o acesso, a reutilização e preservação a longo prazo. De acordo com o DCC<sup>15</sup> os metadados são a espinha dorsal da curadoria digital, sem eles torna-se impossível a recuperação, a interpretação e a utilização. São entendidos como informação descritiva ou contextual que se refere ou está associada a outro objeto ou recurso, assumindo normalmente a forma de um conjunto estruturado de elementos (modelo de metadados) que descrevem o recurso de informação.

Para Day (2005) a definição dos metadados deve estar relacionada com o uso e as funções às quais se destinam apoiar, dessa forma estes podem ter a função de recuperação, compreensão e gestão.

*“Os metadados podem descrever o conteúdo, o contexto e a proveniência dos conjuntos de dados de forma padronizada e estruturada, tipicamente descrevendo o propósito, a origem, as características temporais, a localização geográfica, a autoria, as condições de acesso e os termos de uso de um conjunto de dados.”*(University of Essex 2017a)

Qin, Ball, e Greenberg (2012) acreditam que existe um consenso entre a comunidade científica de que os metadados são a base para a descoberta dos dados, permitem a utilização e a preservação. Isto é possível porque os metadados agregam valor aos conjuntos de dados, esse valor provém da informação adicional que descreve o conteúdo, a proveniência o contexto, e muitos outros aspetos que possibilitam a recuperação, a interpretação e a gestão dos conjuntos de dados.

É frequente a interação com sistemas que usam de forma generalizada os metadados, sem que o utilizador comum se aperceba. Alguns exemplos disso são a pesquisa de vídeos no *Youtube*; ou a lista de contatos do telemóvel. Assim, para que o utilizador possa recuperar o vídeo ou um contato que pretende é necessário que tanto o vídeo, bem como o contato contenham informação que os possam distinguir de outros. A um determinado contato são normalmente associados metadados como o nome, o número, o e-mail, sendo possível associar mais metadados como por exemplo: a morada, a alcunha, o aniversário, a relação entre outros. Por sua vez, aos vídeos podem estar associados metadados como o título; data de publicação; descrição, categoria, tipo de licença, autor e outros.

Os metadados apresentam-se assim como elementos capazes de agregar valor aos dados possibilitando a sua recuperação.

---

<sup>15</sup> DCC. (2018) What are Metadata Standards. [5 de janeiro de 2018] Disponível em: <http://www.dcc.ac.uk/resources/briefing-papers/standards-watch-papers/what-are-metadata-standards>

Os metadados, considerando a sua função, podem ser categorizados em: descritivos; administrativos; técnicos; de preservação e estruturais, sendo que em alguns casos os autores definem mais categorias.

**Metadados descritivos:** descrevem o recurso permitindo melhorar a compreensão e potencializar a sua recuperação (título; autor; conteúdo; data de publicação).

**Metadados administrativos:** informações necessárias para possibilitar a gestão dos recursos e permitir a citação (versão dos dados; formato de arquivo; declaração de direitos).

**Metadados técnicos:** descrevem informação sobre ferramentas necessárias para manipular ou armazenar os dados (tipo de ficheiro; tamanho; data de criação).

**Metadados de preservação:** descrevem processos de gestão, registos de alterações, definem estratégias de preservação futuras.

**Metadados estruturais:** descreve relações entre recursos (lugar na hierarquia).

De acordo com University of Essex (2017b) a garantia de que os dados de investigação podem ser usados, partilhados e reutilizados só é possível se houver uma descrição clara dos mesmos, anotações, informações contextuais, documentação que explique de que forma os dados foram criados, o que significam, quais os conteúdos, estrutura e qualquer manipulação que possam ter sofrido. Documentos como os cadernos de laboratório, protocolos experimentais, dicionários de dados, livros de código, informações sobre a configuração e calibração dos equipamentos, relatórios sobre a metodologia, esquemas sobre o conjunto de dados são componentes fundamentais para apoiarem os gestores no desenvolvimento de metadados.

Segundo o UK Data Archive (2011) para que seja possível utilizar conjuntos de dados é necessária informação suficiente que dê sentido aos dados, dessa forma a atribuição de metadados, referida muitas vezes por documentação, deve incluir informação sobre:

o **contexto de recolha dos dados:** breve explicação sobre o enquadramento, os objetivos e as hipóteses do projeto;

os **métodos de recolha de dados:** descrição dos processos de recolha de dados, quais os instrumentos utilizados para a recolha, definições do hardware e software usados, descrição das amostras, descrição das escalas e resoluções usadas, cobertura temporal e geográfica, outras fontes de dados usadas;

a **definição da estrutura do conjunto de dados:** qual a organização, e relações estabelecidas entre elementos dos conjuntos de dados;

a **definição dos procedimentos de validação de dados:** verificações efetuadas, provas, limpezas e quais as técnicas aplicadas para a garantia da qualidade;

as **alterações efetuadas no conjunto de dados:** registo de todas as alterações efetuadas ao longo do tempo desde a sua criação, identificação de versões;

as **condições de acesso:** definir as condições de acesso, uso, replicação ou confidencialidade dos dados;

as **variáveis:** nomes, rótulos claros para descrever as variáveis, os registros e os valores associados;

a **definição de classificações ou esquemas usados:** explicar códigos, esquemas ou classificações usadas;

a **definição de terminologias:** definir terminologia especializada utilizada ou siglas;

os **dados gerados após a coleta;**

as **variáveis de ponderação.**

De acordo com a análise levada a cabo por Willis Craig, Greenberg e White (2012) existem requisitos para o desenvolvimento de metadados, e estes parecem estar associados a tipos específicos de dados, mas que são aplicáveis em todas as disciplinas. Os requisitos identificados pelos autores são apresentados na Figura 9.

Scheme abstraction	A well-defined metadata scheme will likely outlive its initial rendering. Abstraction allows needs be captured a way that supports multiple renderings over time.
Scheme extensibility, flexibility, and modularity	These are essential design requirements for information systems, including metadata, and will ensure the longevity of the scheme, facilitating adoption and modification over time and extension to meet yet-to-be-identified needs.
Comprehensiveness and sufficiency	Scheme creators should strive to define an element set (or vocabulary) that is comprehensive and also identify a minimal set of elements that are essential for documentation within the domain.
Simplicity	Scheme creators should take into account the levels of technical expertise of their community and support those with minimal as well as those with abundant tools and resources.
Data interchange (exchange)	An essential function of any metadata scheme for scientific data is the ability to exchange, share, and communicate data, whether raw or otherwise, among community members.
Data retrieval	Another essential function of any metadata scheme for scientific data is the ability to discover and acquire the data, taking into account discipline-specific access paths.
Data archiving	An essential function of any repository for the long-term preservation of information. Scheme developers should consider functions defined within the digital preservation community.
Data publication	Research data is an important component of the process of scientific communication and documentation. Scheme developers should account for the association of data with published research, such as citation in peer-reviewed journal articles.

FIGURA 9-REQUISITOS FUNDAMENTAIS PARA A DESCRIÇÃO DE DADOS DE INVESTIGAÇÃO

FONTE: WILLIS CRAIG, GREENBERG JANE (2012)

Assim, tendo em conta as tipologias dos dados é notória a dimensão e complexidade que os modelos de dados devem ter, para conseguirem acrescentar toda a informação adicional necessária aos conjuntos de dados dos vários domínios. O modelo de metadados deve ser desenvolvido de forma a garantir suporte a um conjunto de funções descritivas, administrativas, técnicas e de preservação, dessa forma são especificados elementos, aos quais serão atribuídos valores a quando da descrição dos conjuntos de dados.

Os modelos de metadados são normalmente publicados como documentos de referência aos quais são atribuídas descrições semânticas dos elementos, de modo a garantir que quem irá utilizar o modelo, consegue fazê-lo de forma eficiente. A abordagem comum utilizada para o

desenvolvimento de modelos de metadados é baseada em entidades (pessoas; eventos; objetos; lugares) e nas relações que se estabelecem no domínio (“é parte de algo”; “contém”).

A Research Data Alliance e o grupo de trabalho Metadata Standard Catalogue levaram ao desenvolvimento de um catálogo de modelos de metadados, que pode ser consultado através do DCC.

Uma das várias componentes do DCC é a disponibilização de modelos de metadados disciplinares. Este apresenta uma divisão em cinco áreas que são: a Biologia, a Ciência da Terra, a Ciência Física, as Ciências sociais e humanidades e dados gerais de investigação. Através da análise das cinco divisões decidiu-se seleccionar um modelo de metadados de cada divisão, o DC (Dublin Core), que não foi desenvolvido com o objetivo de tratar dados de investigação, mas ainda assim é usado por várias disciplinas nas descrições de dados. Do domínio da biologia seleccionou-se o EML (Ecological Metadata Language); das ciências da terra o DIF (Directory Interchange Format); da ciência física o CIF (Crystallographic Information Framework) e das ciências sociais e humanidades o DDI (Data Documentation Initiative).

### 2.3.1 DC (Dublin Core)

O modelo de metadados dados Dublin Core <sup>16</sup> (DC) foi desenvolvido pela Dublin Core Metadata Initiative, sendo publicado em 2009 como ISO Standard 15836. É um modelo simples, que descreve objetos digitais. Não foi desenvolvido para nenhum domínio específico, no entanto pela simplicidade e fácil compreensão, é um dos formatos de metadados mais conhecido e amplamente aplicado. O conjunto simples de metadados do Dublin Core são:

**Título:** nome do recurso;

**Autor:** organização, serviço ou pessoa que produziu o conteúdo;

**Assunto:** palavras-chave ou códigos de classificações que sejam representativas do conteúdo;

**Descrição:** descrição sobre o conteúdo;

**Editor:** organização, serviço ou pessoa;

**Contribuidor:** organizações, serviços ou pessoas que contribuíram para o objeto;

**Data:** data da publicação ou disponibilização do recurso;

**Tipo de recurso:** descrição de funções e categorias;

**Formato:** determina o software, hardware necessários para apresentar ou trabalhar o recurso;

**Identificador do recurso;**

**Fonte:** a fonte de que é proveniente o recurso;

---

<sup>16</sup> DC - <http://www.dublincore.org/> acedido em: 19 de janeiro de 2018

**Idioma:** língua padrão do recurso;

**Relação:** com outros recursos;

**Abrangência:** localização espacial, temporal ou de jurisdição;

**Direitos de autor:** descrição dos direitos do recurso;

Os elementos são opcionais, sendo que no momento da descrição dos objetos o responsável deverá selecionar o conjunto que for necessário para descrever o recurso, podendo ainda repetir elementos ou usar em conjunto com outros modelos de metadados. Um exemplo da utilização do DC em conjunto com outros modelos é o British Oceanographic Data Centre Published Data Library (BODC) um centro de dados que trata e distribui dados do ambiente marinho.

### 2.3.2 DDI (Data Documentation Initiative)

A Data Documentation Initiative <sup>17</sup>(DDI), desenvolvida pela DDI Alliance, fornece um padrão internacional para a descrição de dados das ciências sociais, comportamentais e económicas. É usado frequentemente para tratar dados do tipo observacional bem como do tipo experimental.

Atualmente são mantidas duas versões do modelo: o DDI Codebook também conhecido por DDI versão 2 (versão mais simples), já está disponível na versão 2.5; e o DDI Lifecycle ou DDI versão 3 (versão mais desenvolvida que dá suporte à descrição de todo o ciclo de vida dos dados de investigação), disponível também na versão 3.2.

O DDI Lifecycle sugere cinco elementos principais: conceção do estudo; processo de recolha de dados; estrutura lógica de dados codificados; e o arquivo.

Alguns dos elementos das várias versões do DDI são: o modelo de recolha dos dados; o método de agregação; unidade de análise; conjunto de caracteres; tipo de dados; tipos de fontes de dados; tipo de formato dos dados; tipo de evento do ciclo de vida; procedimentos de amostragem; tipo de instrumento; fuso horário entre muitos outros.

Alguns casos de uso do DDI são: o catálogo CESSDA, que fornece uma interface a nível Europeu para dados das ciências sociais, e o UKDA- UK Data Archive a maior organização responsável pela curadoria de dados das ciências sociais e das humanidades do Reino Unido.

---

<sup>17</sup> DDI- <http://www.ddialliance.org/> acedido em: 19 de janeiro de 2018

### 2.3.3 EML (Ecological Metadata Language)

O modelo de dados EML<sup>18</sup>, baseado no trabalho desenvolvido pela Ecological Society of America em conjunto com o Knowledge Network for Biocomplexity, foi pensado para satisfazer as necessidades de descrição dos dados de investigação da disciplina de ecologia. Este modelo serve ainda para descrever domínios da biodiversidade, ecossistema, meteorologia, ciências da terra, entre outras. Assim, é usado frequentemente para descrever dados do tipo experimental e observacional.

O EML é implementado através de um conjunto de documentos XML que permitem descrever os dados a vários níveis, de forma modular (através do uso de estruturas definidas) e extensível (possibilitando a introdução de novos metadados). Cada módulo EML é responsável pela descrição de uma parte dos metadados relativos ao recurso em causa. Um módulo simplificado pode conter o título do recurso; e informação sobre o autor, como por exemplo o nome e o contato.

Um dos vários casos de aplicação do EML é a (GBIF) Global Biodiversity Information Facility, uma organização intergovernamental que facilita o acesso a dados da biodiversidade.

### 2.3.4 DIF (Directory Interchange Format)

O DIF<sup>19</sup> é um formato de dados resultante de uma iniciativa a Earth Science and Applications Data Systems Workshop (ESADS) realizada em 1987. O modelo foi desenvolvido com o objetivo de descrever conjuntos de dados científicos das ciências da terra.

O DIF considera elementos que recolhem informação sobre os instrumentos utilizados na captura de dados; origem; responsáveis pela partilha dos dados, restrições de acesso; características temporais e espaciais; projetos que se relacionam com o conjunto de dados entre outras.

O DIF identifica oito dos seus elementos como obrigatórios, dezassete são altamente recomendados e onze são recomendados. Os oito elementos obrigatórios são: ID de entrada (identificador do conjunto de dados); Título de entrada; Palavras-chave; Categoria do tópico ISO (identifica as palavras-chave na ISO-19115); Centro de dados (identificação dos responsáveis pela distribuição dos dados); Sumário; Nome dos metadados e a versão.

Um exemplo da aplicação do DIF é o NASA EOSDIS The Earth Observing System Data and Information System, fornece dados de investigação, online, de várias fontes desde satélites, aeronaves, entre muitos outros.

---

<sup>18</sup> EML- <https://knb.ecoinformatics.org> acedido em: 22 de janeiro de 2018

<sup>19</sup> DIF- <https://gcmd.nasa.gov/add/difguide/> acedido em: 22 de janeiro de 2018

### 2.3.5 CIF (Crystallographic Information Framework)

O CIF<sup>20</sup> é um modelo de dados bem estabelecido para o tratamento, distribuição e arquivo de recursos resultantes das investigações em cristalografia. Foi desenvolvido pelo grupo Working Party on Crystallographic Information da IUCr (International Union of Crystallography) e patrocinado pela IUCr Commission on Crystallographic Data e por IUCr Commission on Journals, sendo adotado em 1990.

Existem duas versões do CIF versão 1.1 e versão 2.0, a primeira foi desenvolvida no início dos anos 90 sendo que a versão 2.0 aprovada em 2014, veio complementar a versão inicial.

O CIF é composto por um conjunto amplo de descritores, que se ajustam às necessidades de descrição de recursos complexos produzidos na cristalografia. É aplicado essencialmente na descrição de dados do tipo experimental.

O CIF divide os seus descritores em vários conjuntos por exemplo: o Core dictionary (coreCIF); Powder dictionary (pdCIF); Modulated and composite structures dictionary (msCIF), entre outros.

Um exemplo do uso do modelo de dados CIF é a Cambridge Structural Database<sup>21</sup> (CSD), um repositório mundial de estruturas de cristais orgânicos, metálicos e de pequenas moléculas. Foi fundado em 1965 e conta com mais de 900000 entradas de análises de difração de raios-x e neutrões.

O DCC contempla uma secção com modelos de metadados gerais de investigação, ou seja, para disciplinas em que ainda não foram estabelecidos modelos de metadados para o tratamento dos dados de investigação. Assim esta secção apresenta um conjunto de modelos mais amplos, como é o caso do Dublin Core, que servem para tentar descrever mesmo a um nível reduzido os dados de investigação. Segundo Qin, Ball e Greenberg (2012) há evidências de que apesar do esforço para o desenvolvimento e aplicação dos modelos de metadados, a utilização dos mesmos não acompanhou o crescimento dos dados científicos.

De acordo com Thanos (2017) os investigadores devem saber quem tem os dados de que necessitam ou onde os podem encontrar, assim fomentar um ambiente multidisciplinar científico em rede de forma a recuperar dados de investigações relevantes é um grande desafio. Contudo a possibilidade de recuperação dos dados prende-se com a capacidade de descrever adequadamente os dados. Ainda de acordo com Thanos (2017), apesar da importância dos dados de investigação, não é fácil usar os mesmos uma vez que existem vários obstáculos, alguns deles são: a dificuldade de representar dados heterogêneos; existência de uma diversidade de

---

<sup>20</sup> CIF- <http://www.iucr.org/resources/cif> acedido em: 23 de janeiro de 2018

<sup>21</sup> CSD- <https://www.ccdc.cam.ac.uk/solutions/csd-system/components/csd/> acedido em: 23 de janeiro de 2018



consultas de pessoas que falam idiomas distintos; falta de convenções de comunicação entre sistemas; ausência de vocabulários comuns e de terminologias de domínio.

O grande problema é a inexistência de modelos de metadados que representem todas as informações necessárias para os diversos domínios da ciência. Em alguns casos, como o da cristalografia, já existem soluções, como é exemplo o CIF, contudo existem vários domínios com carência de modelos de dados. Constatando a diversidade de dados resultantes das várias disciplinas com origens, processamento, métodos, formatos diversificados é notório que um modelo não é suficiente para aplicar de forma generalizada na descrição dos conjuntos de dados. Um modelo de metadados desenvolvido especificamente para satisfazer todas as necessidades de descrição de um domínio pode tornar-se tão complexo que dificilmente poderá ser aplicado na íntegra num outro domínio. No entanto, é possível que alguns dos elementos de um modelo possam ter interesse para a descrição de outros domínios.

Desta forma, a solução poderá passar pela utilização de ferramentas que disponibilizem vários modelos de metadados que possam ser usados em simultâneo para a descrição dos dados. Para isso, segundo Willis Craig, Greenberg, White (2012) de forma a que os metadados possam ser amplamente aplicados em vários domínios é necessário o desenvolvimento/identificação de requisitos universais.

## 2.4 Sumário

A crescente valorização da ciência aberta amplamente defendida pelas agências de financiamento apresenta-se como um dos grandes desafios para a gestão de dados de investigação, uma vez que necessitam desenvolver soluções que sustentem a partilha de dados de investigação. A reutilização dos dados está dependente do enriquecimento do ponto de vista semântico (Thanos 2017) ou seja o desenvolvimento de metadados é a atividade fundamental que garante a recuperação, acesso e interpretação dos dados.

Enquanto que para as *big sciences* se desenvolveram ferramentas ajustadas às necessidades de descrição para as pequenas comunidades científicas a realidade não se assemelha, sendo nítida a escassez de recursos.

Acredita-se que com o desenvolvimento do conhecimento relativo aos requisitos necessários para a descrição dos diversos domínios é possível auxiliar os investigadores na descrição dos seus conjuntos de dados. Nesse sentido, com este trabalho é proposta a análise de conteúdo baseada em documentos, produzidos por investigadores, que relatem experiências que geram dados do tipo experimental. Assim, serão efetuadas as análises para casos de estudo como a química, a energia renovável, engenharia de materiais e a física de forma a identificar requisitos de metadados necessários para a descrição desses domínios.

### 3 Metodologia

Para o desenvolvimento da dissertação a abordagem metodológica adotada foi o método qualitativo de investigação, mais especificamente a pesquisa exploratória. Esta metodologia permite que o investigador adquira conhecimento sobre o caso de estudo aumentando assim a compreensão sobre os temas que irá abordar. Neste caso específico em que o conhecimento sobre os requisitos necessários para a descrição de metadados do tipo experimental ainda não está muito desenvolvido, a pesquisa exploratória permite o estabelecimento de hipóteses que auxiliem o desenvolvimento do estudo.

Assim, uma das hipóteses que se estabelece é a possibilidade de os requisitos necessários para a descrição de conjuntos de dados terem características diferentes que se relacionam com cada tipologia de dados. O foco desta dissertação serão apenas os dados do tipo experimental, uma vez que o tempo para o desenvolvimento da dissertação é reduzido face ao tempo necessário para tratar todas as tipologias de dados. A razão por se optar pelos dados experimentais relaciona-se com o facto de quando bem documentados podem ser reproduzidos.

A parte prática da dissertação inicia-se com a seleção dos casos de uso de interesse para a investigação. A escolha dos casos de uso é orientada pelas relações estabelecidas com investigadores no âmbito de projetos anteriores, pela facilidade em estabelecer contacto com os mesmos e pela garantia de disponibilidade de investigadores para avaliar as soluções propostas.

O que se propõe é uma abordagem baseada na análise de conteúdo. À semelhança do que defende Chao (2014) a análise de conteúdo produzido por investigadores e publicado, apresenta-se como uma solução alternativa viável que permite extrair informação para o desenvolvimento de metadados. Seguindo a mesma linha de pensamento, foram selecionados e analisados um conjunto de artigos que descrevem experiências semelhantes desenvolvidas nos domínios escolhidos (química sustentável, energias renováveis, engenharia de materiais e física), de forma a compreender quais os conceitos fundamentais que necessitam de ser descritos.

Dos artigos selecionados são analisadas, pormenorizadamente, algumas secções nomeadamente a *Introdução*, e secções que relatam as configurações experimentais das experiências e a metodologia. A opção por algumas das secções torna a análise de conteúdo um processo realista e viável. Pretende-se compreender se o curador de dados sem conhecimento prévio em domínios específicos pode, de forma autónoma, identificar conceitos fundamentais à descrição de dados.

Exemplificando, o que se propõe nesta abordagem é a identificação de “palavras chave” que sirvam de referência para identificar um descritor. Desse modo, sempre que um autor sente necessidade de referir qual é a área superficial do adsorvente “*Activated carbon from SD fine chemicals (surface area 735.60m<sup>2</sup> g<sup>-1</sup>)*” ou registar qual a solução de controlo que utilizou “it

*was controlled using 0.1M NaOH or 0.1M H2SO4 solutions*”, sugere ao curador de dados que descritores possíveis são “Área da superfície do adsorvente” e “Solução de Controle”.

Depois de definido o modelo de metadados, os resultados foram validados com investigadores dos domínios analisados. O esforço que se pede ao investigador é grande uma vez que o processo de investigação é prioritário e este tende a não ter disponibilidade para as questões da gestão dos dados. Dessa forma, pretende-se com esta metodologia limitar a intervenção do investigador ao menor envolvimento possível no momento inicial, sendo a intervenção fundamental na fase de avaliação, onde os investigadores podem contribuir indicando as suas preferências, metadados que não estejam previstos, entre outras sugestões.

No momento da avaliação dos resultados obtidos junto do investigador a abordagem utilizada foi variada. Foram efetuadas sessões de avaliação presenciais; outro caso foi a interação com o investigador de forma remota através de um documento partilhado no Google Drive, onde a intervenção do curador é reduzida (tal como responder a questões colocadas pelo investigador), e ainda uma sessão presencial, mas apresentando numa fase inicial um conjunto de descritores selecionados do modelo de metadados de forma aleatória. O objetivo da aplicação das diferentes abordagens relaciona-se com a tentativa de controlar possíveis condicionantes envolvidas, tentar estimular a interação do investigador, controlar a interação do curador de modo a não influenciar as decisões do investigador.

A argumentação que justifica a necessidade de utilizar diferentes abordagens na interação com o investigador tem a ver tanto com a disponibilidade do investigador bem como com a diferenciação de conhecimentos individuais de cada pessoa, sensibilidades distintas para a gestão de dados de investigação e, por essa razão, uma abordagem rígida não seria ajustada. De acordo com Awre et al. (2015) a gestão de dados de investigação é um problema complicado, e como tal não pode ser tratado como um problema comum. É por isso, necessário ajustar o modo de atuar, uma vez que um pensamento linear não terá sucesso neste contexto. Dessa forma, se os resultados obtidos forem semelhantes nas diferentes formas de abordagem ao longo dos casos de uso a abordagem utilizada poderá ser aceite ou refutada.

## 4 Casos de Uso

Nas secções que se seguem são apresentados cinco casos de uso, tendo sido aplicada a cada um a metodologia descrita anteriormente.

Para cada caso de uso foi analisado um conjunto de artigos e dessa análise resulta uma tabela com expressões e os respetivos descritores. Sempre que possível os modelos de metadados foram validados junto de investigadores do domínio. Em alternativa foi efetuada uma avaliação comparativa com resultados obtidos numa investigação feita por curadores. A Figura 10 resume o processo que deu origem aos casos de uso.

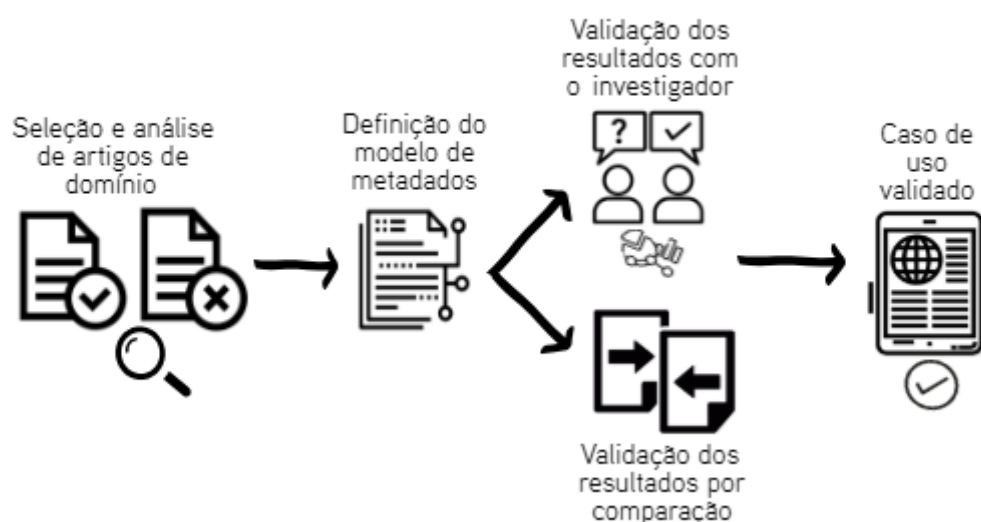


FIGURA 10-PROCESSO QUE ORIGINA OS CASOS DE USO  
FONTE: AUTORIA PRÓPRIA (FERRAMENTA UTILIZADA DISPONÍVEL: [HTTPS://PIKTOCHART.COM/](https://PIKTOCHART.COM/))

Da interação com o investigador, ou da comparação de resultados com outros documentos resulta uma tabela de resultados onde se pode consultar os descritores propostos e a respetiva avaliação. Os descritores que não são considerados corretos estão marcados com a cor laranja. Nos casos em que os descritores foram validados, quando não há informação adicional significa que o descritor foi validado mas não gerou nenhum tipo de anotação, comentário ou sugestão de melhoria.

Os casos de uso que se seguem estão organizados de forma semelhante iniciando com uma breve descrição do domínio em causa, de modo a tentar contextualizar o leitor, seguindo-se a respetiva análise de conteúdo e apresentação dos resultados obtidos.

## 4.1 Química sustentável - Degradação de partículas poluentes

Da atividade humana resultam resíduos que se vão acumulando no meio ambiente, tal como o caso dos desperdícios resultantes dos processos industriais que, ao contaminar a água, o solo e a atmosfera, desencadeiam uma série de problemas para todas as espécies, inclusive a humana. As alterações climáticas, a escassez de água potável, o aparecimento de problemas de saúde, malformações, são algumas das problemáticas que resultam ou que são ampliadas pela poluição.

Numa tentativa de diminuir as repercussões causadas pelos agentes poluentes, é necessário eliminá-los ou torná-los menos ofensivos para o meio ambiente, dessa forma o desenvolvimento de soluções capazes de remover, ou transformar as partículas poluentes tem sido alvo de estudo.

### Análise de conteúdo e resultados

A análise deste domínio teve por base artigos da área que resultam de experiências que estudam compostos químicos através de vários processos que transformam, degradam ou eliminam o poluente, de modo a que este não cause tantos danos ao meio ambiente. A tabela resultante da análise dos artigos pode ser consultada no Anexo I.

A interação com o investigador foi presencial sendo apresentado um conjunto de 60 descritores dos quais 53 foram compreendidos. Entre estes, 38 foram anotados ou aprovados, enquanto para os restantes 15 o investigador sugeriu melhorias. Por outro lado, 7 dos descritores apresentados foram tidos como ambíguos, uns porque não foram compreendidos, outros porque repetem conceitos ou ainda porque não fazem sentido para o investigador. No Anexo II é possível consultar os resultados obtidos através da interação com o investigador.

TABELA I -RESUMO DE RESULTADOS OBTIDOS PARA A QUÍMICA SUSTENTÁVEL: DEGRADAÇÃO DE PARTÍCULAS POLUENTES

Descritores compreendidos		Exemplos
Anotados/aprovados	38	<i>Tamanho da cristalite da amostra: 10nm; Agente oxidante: Oxigénio; Quantidade da amostra centrifugada: 50ml Instrumento de análise dos catalisadores; Instrumento de medição da radiação eletromagnética.</i>
Com sugestão de melhoria	15	<i>Composto químico → Amostra; Área interfacial → Distância interfacial;</i>

		<i>Taxa de fluxo volumétrico na fase gasosa → Fluxo da fase gasosa; Teor de impurezas → Grau de pureza.</i>
Total	53	
Descritores ambíguos		Exemplos
Não compreendidos	3	<i>Instrumento de medição espectral; Absorvência; Concentração interfacial de ozono.</i>
Repetem conceitos	3	<i>Velocidade superficial do gás; Taxa de fluxo de massa de ozono; Quantidade da amostra centrifugada diluída.</i>
Não fazem sentido	1	<i>Comprimento de onda do catalisador.</i>
Total	7	

O investigador afirma que sente necessidade de efetuar registos das experiências e que esses registos são efetuados em atas, que funcionam como um diário onde são anotadas condições de trabalho, condições em que se desenvolveram as experiências ou acontecimentos inesperados. Quando se questionou o investigador se o modelo de metadados apresentado era suficiente para descrever o domínio em questão, e se capta a informação que regista nas atas, a resposta foi positiva, uma vez que o modelo apresentado foi capaz de captar a informação que anota, os descritores fazem sentido. Contudo, contém mais elementos do que necessita para descrever a sua experiência.

Por sua vez, questionado sobre o que considerava ser necessário acrescentar ao modelo de metadados, o investigador sugere que especificava algumas concentrações como exemplo *Concentração da solução*. Na opinião do mesmo, este elemento poderia substituir o elemento *Solução* uma vez que ao anotar a concentração da solução os investigadores iriam anotar também dentro do descritor *Concentração da solução* qual era a solução em causa.

Por fim, o investigador sugeriu, para tornar o processo do curador mais eficiente, a consulta da informação dos instrumentos utilizados nas investigações, uma vez que esses têm campos predefinidos. Acrescenta que muito do trabalho que desenvolvem atualmente passa pela introdução de valores nos instrumentos que devolvem cálculos, ou resultados. Assim, se o curador de dados tivesse oportunidade de aceder às configurações dos instrumentos poderia, com alguma facilidade, encontrar a informação de que necessita para desenvolver o modelo de metadados. No entanto, para que isso fosse possível era necessário aceder aos grupos de investigação, sendo que isso foi logo identificado como uma grande limitação. Mesmo para os investigadores da área, o acesso aos grupos de investigação é extremamente difícil “é muito difícil

aceder, para montar um caso de estudo era necessário pedir autorização a várias pessoas ligadas aos laboratórios” (Afirmação do investigador).

Do estudo efetuado verifica-se a existência de vários requisitos que são necessários descrever para contextualizar as experiências, assim identificam-se os seguintes:

TABELA 2-REQUISITOS NECESSÁRIOS PARA A DESCRIÇÃO DA QUÍMICA SUSTENTÁVEL: DEGRADAÇÃO DE PARTÍCULAS POLUENTES

<b>Categoria</b>	<b>Descritores</b>
Propriedades amostra	Amostra / Elemento químico / Quantidade do composto degradado / Tamanho da cristalite da amostra / Volume do poro da amostra / Distância interfacial / Gás em estudo / Fluxo da fase gasosa / Carbono inorgânico / Fórmula do polifenol / Massa do polifenol / Agente oxidante / Potencial de oxidação / Massa molecular do polifenol / Massa molecular do gás / Reagentes / Grau de pureza / Pressão parcial de ozono na fase gasosa / Carbono orgânico total / Absorbância / Carbono total / pH da solução / Tamanho da partícula / Adsorvente / Área da superfície do adsorvente / Teor de cinzas do adsorvente / Tamanho da partícula do adsorvente / Fórmula molecular do adsorvente / Amostra de referência / Solução de controlo / Solução de limpeza / Solução ácida
Instrumento, características e calibrações	Reator de ozonização / Instrumento de medição da absorbância / Instrumento de radiação de luz UV / Instrumento de medição da intensidade da luz / Instrumento de medição do pH / Instrumento de análise dos catalisadores / Instrumento de medição da radiação eletromagnética / Instrumento para a medição da área superficial / Vaso de reação fotocatalítica / Quantidade da amostra centrifugada
Tempo	Tempo de ozonização / Tempo de medição da intensidade da luz / Tempo de agitação da suspensão
Temperatura	Temperatura de secagem da amostra
Método	Método de análise
Técnica	Técnica de remoção de partículas / Técnica de medição da área superficial
Outros	Condições atmosféricas / Atividade fotocatalítica / Intensidade da luz solar

O investigador necessita definir qual é a amostra que está a investigar e as suas propriedades. É importante referir que a amostra pode incluir simultaneamente o poluente, o catalisador, o adsorvente, o agente oxidante, entre outros, sendo necessário captar informação sobre cada um deles.

Assim, parecia preponderante apresentar ao investigador descritores como: *Potencial de oxidação do agente oxidante, Tamanho da partícula do poluente, Tamanho da partícula do catalisador, Tamanho da partícula do adsorvente, Fórmula molecular do adsorvente*. No entanto, no momento em que o investigador se deparou com alguns destes descritores sugeriu para *Tamanho da partícula do poluente* e *Tamanho da partícula do catalisador*, fosse apenas *Tamanho da partícula*. Para

*Potencial de oxidação do agente oxidante* também sugeriu apenas *Potencial de oxidação*. Por outro lado, para os descritores: *Tamanho da partícula do adsorvente* e *Fórmula molecular do adsorvente* o investigador anotou os valores 80 nm e carbono, respetivamente, sem efetuar nenhuma sugestão.

Também no caso da categoria Instrumento, características e calibrações foram apresentados descritores: *Instrumento* e instrumentos com funções específicas como por exemplo, *Instrumento de medição da absorbância*, *Instrumento de radiação de luz UV* para este caso o investigador referiu que quando o curador de dados especifica a função do instrumento é mais fácil descrever. Este acontecimento levanta algumas questões, por exemplo, se um nível mais geral é suficiente para captar a informação necessária sem que se perca informação relevante. O que poderá depender da disponibilidade e sensibilidade do investigador para descrever os seus dados. Por outro lado, se se considerar especificar demasiado os descritores, o modelo de metadados pode tornar-se demasiado extenso e consequentemente impraticável. Um estudo mais aprofundado terá de ser desenvolvido para dar resposta a estas questões.

Importa ainda registar todos os instrumentos utilizados ao longo da experiência por exemplo, *Reator de ozonização*, *Instrumento de medição do pH*. Note-se que para alguns instrumentos é fundamental recolher informação sobre calibrações específicas que foram efetuadas. Contudo, para este caso, não foi possível identificar um descritor que captasse essa informação, apesar de nos artigos analisados por vezes referirem aspetos específicos de calibrações. É igualmente importante sempre que possível recolher informação sobre características específicas associadas à utilização de um instrumento que possam influenciar a experiência, como exemplo o tipo de lâmpada que um determinado instrumento tem, assim como a quantidade da amostra centrifugada que está dependente da utilização de um instrumento de centrifugação.

Recolher informação sobre o tempo despendido a realizar determinada tarefa também é um aspeto de grande importância, uma vez que se um investigador medir a intensidade da luz durante 10 minutos, ou variar o tempo de medição para 50 minutos os dados registados serão diferentes.

Importa ainda que o investigador registe informação sobre a temperatura, os métodos e as técnicas aplicadas ao longo da experiência uma vez que dependendo da técnica, método aplicado ou da temperatura registada os valores obtidos através da experiência poderão variar.

## **4.2 Energias Renováveis - Produção de camadas absorventes - Filmes finos**

A energia solar é uma das várias fontes de energia desenvolvida para diminuir o uso e as repercussões das energias poluentes. A energia fotovoltaica é um método desenvolvido para



converter a luz solar em energia elétrica. Esta transformação ocorre através de materiais semicondutores que têm a particularidade de libertar energia quando excitados pela luz solar.

Devido à demanda de soluções não poluentes, que garantam o fornecimento de energia elétrica sustentável, têm se desenvolvido vários estudos, como o estudo de CuInS<sub>2</sub> and CuInGaS<sub>2</sub> baseados em filmes finos ou ainda estudos sobre as propriedades óticas do cobre, gálio, selênio entre outros.

### Análise de conteúdo e resultados

O conjunto de documentos de domínios analisados resultam de experiências que avaliam as propriedades das células solares de “thin films” de componentes específicos, com o objetivo de aplicar em painéis fotovoltaicos, melhorando assim a capacidade de produção de energia, a tabela resultante desta análise pode ser consultada no Anexo III.

A interação com o investigador foi efetuada de forma remota, uma vez que este não se encontrava no país. Assim a secção de avaliação foi efetuada através de um documento partilhado no Google Drive.

Da análise efetuada resultou um conjunto de 56 descritores apresentados ao investigador dos quais 45 foram compreendidos, destes 43 foram anotados e 2 obtiveram sugestões de melhoria, os 11 restantes foram considerados ambíguos.

Da interação com o investigador resultou uma tabela que apresenta todos os descritores obtidos da análise inicial e a respetiva validação efetuada pelo investigador, essa tabela está disponível no Anexo IV.

TABELA 3-RESUMO DE RESULTADOS OBTIDOS PARA ENERGIAS RENOVÁVEIS: PRODUÇÃO DE CAMADAS ABSORVENTES: FILMES FINOS

Descritores compreendidos		Exemplos
Anotados/aprovados	43	<p>Composto químico: CuInGaS thin film, CIS, CZTS, ZnO;</p> <p>Coeficiente de absorção: <math>10^5</math></p> <p>Rendimento do composto: 19%;</p> <p>Instrumento: Potentiostat</p> <p>Tempo de secagem da amostra: 10 min;</p> <p>Temperatura de recozimento: 370°C;</p> <p>Técnica de caracterização: XRD, TEM, MEB, UV- visible</p>
Com sugestão de melhoria	2	<p>Transmitância ótica → Transmitância</p> <p>Constante dielétrica → Parte real da constante dielétrica e Parte imaginária da constante dielétrica;</p>

Total	45	
Descritores ambíguos		Exemplos
Não fazem sentido	11	<i>Viscosidade do composto;</i> <i>Ponto de ebulição do composto;</i> <i>Tempo de pulverização;</i> <i>Reagentes;</i> <i>Resistência elétrica;</i> <i>Matriz da solução;</i> <i>Energia do fóton;</i> <i>Fonte de pulverização catódica;</i> <i>Frequência de rádio;</i> <i>Potência da amostra;</i> <i>Tempo de estabilização da temperatura;</i>
Total	11	

O investigador reconhece a necessidade de efetuar o registo das configurações das experiências admitindo que o faz normalmente num caderno, ou ficheiro de texto. Quando questionado sobre a existência de similaridade entre os descritores que lhe foram apresentados e os registos que efetua habitualmente, o investigador respondeu afirmativamente, os conceitos são semelhantes. No entanto, refere que sente a necessidade de anotar também as concentrações dos precursores e no modelo de metadados apresentado essa necessidade não era satisfeita. Por essa razão defende que são necessários mais metadados para captar toda a informação que necessita registar, ainda assim, refere que os parâmetros contextuais são maioritariamente captados.

Da análise efetuada verifica-se a existência de vários requisitos necessários para a descrição do contexto experimental. Na Tabela 4 é possível consultar os descritores identificados, notando que, tal como a investigadora referiu “são necessários mais”.

TABELA 4-REQUISITOS NECESSÁRIOS PARA A DESCRIÇÃO DAS ENERGIAS RENOVÁVEIS: PRODUÇÃO DE CAMADAS ABSORVENTES; FILMES FINOS

Categoria	Descritores
Propriedades amostra	Elemento químico /Composto químico/Intervalo da banda/ Coeficiente de absorção/ Rendimento do composto/ Estado físico do composto/ Potencial de deposição/ Tipo de semiconductor/ Potencial de variação/ Agente complexante/ Solvente/ Solute/ Tipo de reação/ Composição do banho/ Eléctrodo de trabalho/ Eléctrodo de referência/ Contra eléctrodo/ Potencial catódico/ Camada absorvente/ Transmitância/ Comprimento de onda/ Secagem da amostra/ Tipo de lingote/ Tipo de substrato/ Dimensão do substrato/ Tipo de limpeza do substrato/ Absorbância

Instrumento, características e calibrações	Instrumento
Tempo	Tempo de secagem da amostra/ Tempo de deposição/ Tempo de limpeza/ Tempo de recozimento
Temperatura	Temperatura da atmosfera/ Temperatura de secagem da amostra/ Temperatura de recozimento/ Temperatura do substrato
Método	Método
Técnica	Técnica de produção de camadas absorventes/ Técnica de caracterização
Outros	Condições de pressão/ Condições atmosféricas/ Condições de temperatura/ Energia do intervalo/ Índice de refração/ Coeficiente de extinção/ Parte real da constante dielétrica/ Parte imaginária da constante dielétrica

O desenvolvimento deste tipo de experiência engloba vários métodos, técnicas e instrumentos. Desta forma quando um investigador opta por determinada técnica, método ou instrumento em detrimento de outro, irá influenciar alguns dos elementos necessários para efetuar a descrição. Descritores que captem valores da *Temperatura de recozimento*, do *Agente complexante*, do *Tempo de deposição* só serão necessários para técnicas que sujeitem as suas amostras a este tipo de fatores. Por essa razão alguns dos requisitos identificados poderão não ser úteis para técnicas que não considerem esse tipo de aspetos.

Para os descritores *Transmitância ótica* o investigador sugere que deveria ser apenas *Transmitância* e para *Constante dielétrica* sugere uma subdivisão em dois descritores *Parte real da constante dielétrica* e *Parte imaginária da constante dielétrica*.

Da análise efetuada parece relevante que sejam descritas as propriedades da amostra como por exemplo os *Elementos químicos* ou *Compostos químicos* utilizados, ou o *Coeficiente de Absorção*. Além dos que foram apresentados, o investigador sugere que é importante contemplar um descritor para capturar a propriedade da amostra relativamente à capacidade de absorção sugerindo o descritor *Absorbância*.

A descrição de todos os instrumentos utilizados é algo de grande relevância, no entanto não foi possível especificar os diversos instrumentos as características e calibrações, sendo apresentado, ao investigador apenas o descritor *Instrumento*. Este foi anotado sem qualquer tipo de menção relativamente a uma necessidade de maior especificação, o mesmo não se verificou no Caso de Uso da Química Sustentável.

É necessário registar o tempo e as temperaturas de acontecimentos ou fases importantes que tenham influência nos dados produzidos por exemplo *Tempo de secagem da amostra* ou a *Temperatura de recozimento*.

É importante registar também quais as condições a que foram sujeitas as amostras como *Condições de pressão* e as *Condições de temperatura* a que se desenvolveram as experiências. Além destes foi apresentado ao investigador o descritor *Condições atmosféricas*, para captar

simultaneamente valores de pressão e de temperatura. Contudo o investigador anotou este descritor sem fazer qualquer tipo de referência.

Estes são apenas alguns dos elementos que são necessários considerar quando o curador de dados se depara com o domínio em questão. Contudo, tal como foi referido anteriormente muitos dos descritores necessários dependem das técnicas, dos métodos e dos instrumentos selecionados ao longo do desenvolvimento da experiência, sendo este modelo de metadados apenas um ponto de partida.

### **4.3 Física - Síntese de nanopartículas**

O desenvolvimento de estudos sobre nanopartículas é de grande relevância dada a “ampla gama de aplicações” (Valverde-Alva et al. 2015) possíveis em vários domínios. As oportunidades de investigação neste campo são muito ricas dada a capacidade de manipular os materiais a uma escala nano, revolucionando as características e funcionalidades dos materiais.

As aplicações são variadas, na construção podem ter apenas como objetivo melhorar as condições dos produtos através do aumento da duração dos materiais, ou então a adição de novas funcionalidades. Por exemplo, a aplicação de nanopartículas que possibilitam que os materiais sejam capazes de se manterem limpos apresentando a capacidade se limparem autonomamente. As nanoestruturas possibilitam a interação com células específicas que anteriormente não era possível dada a sua dimensão, assim é praticável aplicar esta tecnologia no diagnóstico, bem como nos cuidados de saúde, como por exemplo em transplantes ou engenharia de tecidos. São ainda exemplos a aplicação de nanoestruturas no âmbito da energia, na tentativa de desenvolver um sistema de energia mais sustentável, ou ainda aplicações na indústria automóvel como são exemplo os materiais anti riscos.

#### **Análise de conteúdo e resultados**

A análise do domínio foi baseada no estudo de artigos que relatam experiências que estudam a síntese de nanopartículas por laser pulsado. Os resultados obtidos nesta fase podem ser consultados no Anexo V.

Numa perspetiva de estimular a interação do investigador e gerar discussão foi apresentada uma versão simplificada dos descritores identificados. A seleção foi meramente aleatória, selecionando de um conjunto de 74 descritores apenas 23 para apresentar numa fase inicial.

Dos 23 descritores apresentados 18 foram anotados, ou avaliados, para 3 descritores a investigadora fez sugestões de melhoria sendo que a investigadora afirmou que apenas dois não faziam sentido. A tabela de resultados pode ser consultada no Anexo VI.

TABELA 5-RESUMO DE RESULTADOS OBTIDOS PARA FÍSICA-SÍNTESE DE NANOPARTÍCULAS

Descritores compreendidos		Exemplos
Anotados/aprovados	18	<p>Amostra: GSG5 – nanopartículas de <math>Gd_5Si_{1.8}Ge_{2.2}</math>;</p> <p>Produtor da amostra: “Eu” (investigador);</p> <p>Revestimento da amostra: Substrato de silício;</p> <p>Dimensões do revestimento 1x2mm;</p> <p>Estabilizador: Atmosfera de argônio;</p> <p>Tamanho da partícula: 30 nm;</p> <p>Solução;</p> <p>Agente redutor;</p> <p>Instrumento de análise das propriedades óticas;</p> <p>Instrumento de síntese da amostra: pulsed laser deposition;</p> <p>Instrumento de emissão de radiação: KrF Excimer;</p> <p>Tempo de duração do pulso: 20 ns;</p> <p>Temperatura de síntese: temperatura ambiente;</p> <p>Temperatura de aquecimento da mistura: 2°C;</p> <p>Método de síntese da amostra: gas phase synthesis;</p> <p>Técnica de caracterização: HR-TEM, XRD e SQUID;</p> <p>Condições atmosféricas: Argon – 1 Torr;</p> <p>Coletor: substrato de Silício;</p>
Com sugestão de melhoria	3	<p>Concentração da amostra → Massa da amostra;</p> <p>Largura do pulso do laser → Energia do pulso</p> <p>Tempo de aquecimento da mistura → Tempo de deposição</p>
Total	21	
Descritores ambíguos		Exemplos
Não fazem sentido	2	<p>Concentração da molécula de passivação</p> <p>Resistência Milli-Q</p>
Total	2	

O investigador afirma que tem uma necessidade de registar o contexto experimental e os resultados que vai obtendo ao longo do processo de investigação, que “não consegue trabalhar no caos” e que se disciplina de forma a registar toda a informação de que necessita. Para tal, usa habitualmente o PowerPoint porque além de permitir registar todo o processo, têm sempre preparada uma apresentação sobre a investigação, caso necessite. Refere que tentou usar outros métodos, como registo de um diário, ou de cadernos, mas que não consegue organizar-se de modo tão eficiente.

Pretendia-se com a seleção de um conjunto de descritores, escolhidos aleatoriamente, tentar estimular a interação do investigador verificando se sentiria necessidade de mais descritores, ou se ficaria surpreso ao verificar que lhe eram apresentados apenas três

instrumentos muito específicos. Exemplificando, apresentou-se ao investigador apenas *Instrumento de síntese da amostra*, *Instrumento de emissão de radiação* e *Instrumento de análise das propriedades óticas* de um conjunto de 20 instrumentos identificados. Acreditava-se que poderiam ser poucos tendo em conta o conjunto total, que o investigador necessitaria de mais, ou que estes não fossem úteis para o seu contexto experimental. No entanto, a investigadora anotou os dois instrumentos sem fazer qualquer tipo de pergunta, ao que se questionou se teria preferência por um descritor Instrumento respondeu que prefere que os instrumentos estejam especificados. Sugerindo que poderia ser considerado um campo que descrevesse o produtor do instrumento, que em algumas investigações esse aspeto poderia ser relevante.

Relativamente ao descritor *Revestimento da amostra* sugere que este não é o melhor termo, mas que se compreende e que muitas pessoas o usam habitualmente, não sabendo indicar qual poderia ser o termo substituto.

Para o descritor *Solução* o investigador afirma que será mais importante para a química; relativamente a *agente redutor* compreende o descritor, mas não necessita de o usar; quanto ao *instrumento de análise das propriedades óticas* afirma ter importância para quem lida com questões das propriedades óticas, mas que não é o caso.

No que diz respeito à *Concentração da amostra* a investigadora anotou 0.5mg, no entanto, sugere que Massa da amostra é mais apropriado. Igualmente, concentração da amostra é relevante para outro tipo de experiências. Após reflexão acrescenta que também pode ser importante para a sua experiência em específico no caso dos compósitos, pelo que sugere que se possa manter os dois descritores.

Para a *Largura do pulso do laser* no qual a investigadora anotou “248 nm (comprimento de onda); 500 mJ (energia do pulso); 10 Hz (frequência do pulso); 20 ns (duração do pulso)” sugere que Energia do pulso seria um descritor melhor. De destacar que no conjunto de descritores que até então não tinham sido apresentados ao investigador foi identificado um descritor semelhante à sugestão que é *Energia do laser por pulso*. Estavam ainda previstos os descritores *Comprimento de onda do pulso*, *Frequência do pulso* e *Tempo de duração do pulso*. Sabendo qual o instrumento utilizado, informação captada pelo descritor Instrumento de emissão de radiação a investigadora afirma que não seria necessário anotar o comprimento de onda, uma vez que, ao indicar o valor para esse instrumento (KrF Excimer) identifica-se o comprimento de onda, a frequência e a duração do pulso. Sendo apenas necessário descrever a Energia do pulso uma vez que este valor pode variar de acordo com o que o investigador deseja.

Questionou-se o investigador se faltaram descritores importantes, o investigador respondeu que o conjunto considerava mais informação do que necessitava. Acrescentou que se enviasse os metadados a outro investigador ele iria ser capaz de reproduzir a experiência.

Após a avaliação dos 23 descritores a investigadora mostrou disponibilidade para colaborar. Assim solicitou-se um esforço extra apresentando o modelo de metadados completo de forma a que identificasse quais os conceitos que não eram relevantes.

Dos 74 apenas 3 descritores foram considerados ambíguos, uns porque não faziam sentido, outros porque repetiam conceitos.

TABELA 6-RESUMO DE RESULTADOS OBTIDOS PARA FÍSICA - SÍNTESE DE NANOPARTÍCULAS (DESCRITORES AMBÍGUOS)

Descritores ambíguos		Exemplos
Não fazem sentido	2	<i>Concentração da molécula de passivação</i> <i>Resistência Milli-Q</i>
Repetem conceitos	1	<i>Instrumento de análise por microscopia eletrônica de varredura</i>
Total	3	

Do estudo efetuado verifica-se a necessidade de descrever vários requisitos de modo a contextualizar as experiências, identificaram-se os seguintes:

TABELA 7-REQUISITOS NECESSÁRIOS PARA A DESCRIÇÃO DA FÍSICA - SÍNTESE DE NANOPARTÍCULAS

Categoria	Descritores
Propriedades amostra	Amostra/ Produtor da amostra/ Concentração da amostra/ Massa da amostra/ Dimensões da amostra/ Gás carreador/ Pressão do gás carreador/ Revestimento da amostra/ Dimensões do revestimento/ Pressurização/ Concentração de refluxante/ Molécula de passivação da superfície/ Solução/ Profundidade de solução acima da amostra/ Distância à base do recipiente de sintetização/ Agitação da solução/ Concentração da solução/ Semicondutor/ Dimensões do semicondutor/ Agente redutor/ Estabilizador/ Tamanho da partícula/ Reagente/ Concentração do solvente/ Quantidade de solução
Instrumento, características e calibrações	Instrumento de síntese da amostra/ Recipiente de sintetização da amostra/ Dimensões do recipiente de sintetização da amostra/ Instrumento de emissão de radiação/ Energia do pulso/ Comprimento de onda do pulso/ Comprimento da irradiação/ Distância focal da lente/Frequência do pulso/ Instrumento de análise FESEM/ Instrumento de medição de Difração de Raios X/ Instrumento de registo da espectroscopia/ Instrumento de absorção UV/ VIS/ Instrumento de análise da morfologia / Instrumento de análise das propriedades óticas/ Instrumento de purificação da água/ Instrumento de irradiação de harmónico/ Instrumento de caracterização da estrutura cristalina/ Radiação/ Comprimento de onda/ Taxa de varredura/ Instrumento de remoção do excesso de solução/ Instrumento de análise química/ Instrumento de deteção de vibrações/ Instrumento de medição da energia transmitida/ Instrumento de identificação da concentração da amostra nos coloides

Tempo	Tempo de duração do pulso/ Tempo de deposição/ Tempo de ablação/ Tempo de secagem da amostra
Temperatura	Temperatura de síntese/ Temperatura de aquecimento da mistura/ Temperatura de secagem da amostra
Método	Método de síntese da amostra/ Método de tratamento e passivação da amostra/ Método de remoção da molécula de passivação
Técnica	Técnica de caracterização/ Técnica de preparação de coloides
Outros	Condições atmosféricas/ Produtividade por pulso/ Intervalo da banda/ Área de incisão do feixe/ Coletor

O investigador necessita registrar qual é a amostra da sua investigação e as suas propriedades, necessita indicar quais os instrumentos que utiliza e se necessário registrar também características e calibrações do instrumento que possam influenciar os resultados. Importa também captar informação sobre tempo, temperaturas, métodos e técnicas aplicadas, uma vez que ao variar algum destes fatores os resultados podem ser influenciados.

Sobre a questão, se a abordagem facilita a compreensão do que se pretende como os metadados o investigador respondeu que sim, facilita. No entanto, sugere que a organização dos metadados de forma diferente facilitaria a interação do investigador, afirma que “os elementos estão muito misturados”. Compreendendo também que estabelecer uma organização seria uma tarefa difícil para alguém sem conhecimento do domínio.

A determinado momento da análise dos descritores o investigador conclui que a importância dos descritores irá depender “tudo vai depender do que se usar”. Isto significa que dependendo do método, técnica, amostra, instrumento escolhidos. A seleção dos descritores necessários vai ser influenciada. Por exemplo o descritor *Dimensões do recipiente de sintetização da amostra* só será necessário se na experiência for utilizado um recipiente de sintetização.

#### 4.4 Engenharia Mecânica - Mecânica de fratura: Feixe cantiléver

A aplicação variada dos materiais deve-se em grande parte ao desenvolvimento de estudos que permitam avaliar as suas características e assim verificar qual o desempenho no desenvolvimento das funções a que foram destinados.

Considerando as estruturas construídas pelo homem como pontes, edifícios, navios, aviões é evidente que a segurança das populações está dependente da resistência dos materiais utilizados na construção dessas estruturas. A queda de aviões, construções, têm consequências irreparáveis na vida das pessoas. Neste sentido, o desenvolvimento de estudos sobre os materiais é de extrema importância, uma vez que a “detecção precoce de danos nas estruturas pode aumentar o seu tempo de vida útil” (Dahak, Touat, e Benseddiq 2017).



De acordo com Møberg, Budzik, e Jensen (2017) os testes desenvolvidos em feixe cantiléver (simples ou duplos) são os procedimentos experimentais mais comuns aplicados para avaliar a “tenacidade à fratura interfacial dos materiais”.

## Análise de conteúdo e resultados

A análise do domínio em questão teve por base um conjunto de artigos que relatam experiências sobre a resistência de materiais à fratura. O resultado obtido da análise de conteúdo pode ser consultado no Anexo VII.

Numa tentativa de validar os resultados esperava-se contactar um investigador do domínio, contudo não foi possível. Razão pela qual foi efetuada uma breve avaliação por comparação com um trabalho anteriormente desenvolvido, por um curador de dados, que obteve aprovação dos seus metadados junto de um investigador. Os descritores propostos por Castro, Rocha da Silva, e Ribeiro (2013) só foram analisados depois de se efetuar a análise de conteúdo, de forma a que os resultados não fossem influenciados. Na Figura 11- é possível analisar os resultados obtidos no artigo Designing an Application Profile Using Qualified Dublin Core: A Case Study with Fracture Mechanics Datasets.

Descriptor	Description	Example
eml:methods	Procedures that are used in the creation or the subsequent processing of a dataset	Free text
eml: instrumentation	Description of the instruments used in the data collection or quality control and quality assurance	INSTROM-1125
fm:specimen	Type of specimen used in the experiment	Pinuspinaster (Wood)
fm:specimenLenght	Speciment geometric length	L= 400 mm
fm:specimenWidth	Specimen geometric width	B=20mm
fm:specimenHeight	Specimen geometric height	2h = 20 mm
fm:specimenInitialCrackLenght	The crack in the double cantilever beam specimen prior to the fracture test	ao= 150mm
fm:specimenProperties	Specimen specific properties	Free text
fm:temperature	The ambient temperature of the experiment location	18°C
fm:moisture	The moisture percentage at the experiment location.	55
fm:testVelocity	Velocity at which the sampling machine pressed in the sample during the experiment	3mm/m

FIGURA 11-DESCRIPTORES IDENTIFICADOS PELO CURADOR

FONTE: CASTRO, ROCHA DA SILVA E RIBEIRO (2013)

Da análise de conteúdo efetuada obteve-se um conjunto de 64 descritores dos quais 43 estavam relacionados com os descritores identificados pelos autores referidos anteriormente. Relativamente aos 21 descritores restantes não foi possível estabelecer qualquer tipo de relação

e consecutivamente avaliação. No Anexo VIII é possível consultar os resultados obtidos e a avaliação comparativa dos mesmos.

TABELA 8-RESUMO DE RESULTADOS OBTIDOS PARA ENGENHARIA MECÂNICA - MECÂNICA DE FRATURA: FEIXE CANTILÉVER

<b>Descritores dos autores Castro, Rocha da Silva &amp; Ribeiro</b>	<b>Descritores identificados equivalentes</b>
<i>Espécimen</i>	<i>Amostra/Espécimen</i>
<i>Propriedades do espécimen</i>	<i>Dimensões da amostra; Adesivo da amostra; Amostra de referência; Adesivo da amostra de referência; Fornecedor do adesivo de referência; Largura da camada adesiva; Módulo de Young/elasticidade; Compressão da amostra; Taxa de liberação de energia; Erro de cálculo da taxa de liberação de energia; Força de rendimento; Resistência final; Resistência à compressão; Coeficiente de Poisson; Laminado; Dimensões do laminado; Módulo de Young/elasticidade do laminado; Resistência à tração do laminado; Adesivo; Módulo de Young/elasticidade do adesivo; Resistência à tração do adesivo; Número de ancoras; Dimensões da ancora; Ancora; Força de ancoragem; Fornecedor do adesivo; Altura da camada adesiva; Módulo de cisalhamento; Força de cisalhamento tangente à frente da fratura.</i>
<i>Altura do espécimen</i>	<i>Altura da amostra</i>
<i>Comprimento do espécimen</i>	<i>Comprimento da amostra</i>
<i>Largura do espécimen</i>	<i>Largura da amostra</i>
<i>Instrumento</i>	<i>Instrumento de medição 3D; Instrumento de medição de deformação; Instrumento de medição de tensão; Instrumento de medição de deflexão; Instrumento de teste de falha; Instrumento de captura do campo de deflexão; Tensor de tensão</i>
<i>Método</i>	<i>Método de preparação da amostra</i>
<i>Velocidade do teste</i>	-----
<i>Comprimento da fissura antes da experiência</i>	<i>Comprimento da fissura;</i>
<i>Temperatura</i>	<i>Condições atmosféricas</i>
<i>Humidade</i>	
<b>Total: 11</b>	<b>Total: 44</b>
<b>Total de descritores que não estabeleceram relação: 21</b>	
<i>Tempo de observação; Tempo de cura do adesivo; Técnica de detecção de dano; Teste de avaliação da resistência à fratura; Teste de tensão de deformação; Força externa aplicada; Energia da fratura; Fator de intensidade de stress; Raio da curvatura longitudinal; Raio de curvatura transversal; Tensão residual; Deslocamento da fenda; Tipo de falha; Ductilidade; Tensão; Distância entre a ancora e a extremidade de compressão; Profundidade da fratura; Faixa de medição válida; Deslocamento transversal; Largura da fissura; Localização da fissura</i>	

Os autores Castro, Rocha da Silva, e Ribeiro, 2013 identificaram um descritor capaz de captar a amostra utilizada no desenvolvimento da experiência, definindo para isso o descritor *Espécimen*. Igualmente, da análise de conteúdo efetuada, resulta *Amostra/Espécimen*, a sugestão recai sobre estes dois descritores para possibilitar selecionar o termo preferencial do investigador, para o domínio em causa. Observou-se através da investigação dos autores anteriores que o descritor eleito é *Espécimen*, nesse sentido todos os descritores que mencionam a amostra devem ser referidos como espécimen. Exemplificando, *Amostra de referência* deverá ser preferencialmente *Espécimen de referência*.

Relativamente ao descritor *Propriedades do espécimen* é definido pelos autores, um descritor mais geral que possibilita ao investigador descrever livremente o que considerar relevante. Isto pode ser extremamente útil quando não se consegue especificar todas as propriedades da amostra, dando assim flexibilidade para que o investigador descreva o que considerar importante. Por outro lado, a descrição fica dependente da sensibilidade do investigador para a gestão dos dados. Neste caso, através da análise de conteúdo desenvolvida foi possível definir propriedades específicas que se relacionam com a amostra como exemplo, *Amostra de referência*; *Adesivo da amostra de referência*; *Módulo de Young/elasticidade*; *Compressão da amostra*, entre outros que se pode consultar na Tabela 8.

Note-se que os descritores propostos não foram analisados individualmente por um investigador pelo que pode existir alguma inconsistência, necessitando de se verificar se efetivamente todos os descritores dizem respeito a propriedades relacionadas com o espécimen. A especificação das propriedades da amostra orienta o investigador, no entanto, podem faltar propriedades que o investigador necessite e não estejam previstas. Encontrar uma forma adequada e exequível é difícil pelo que se tem de compreender qual a melhor solução. Acredita-se que neste caso, a par das propriedades específicas poderia ser definido também um descritor mais geral como *Propriedades do espécimen* para captar informação que não esteja prevista.

Tal como Castro, Ribeiro da Silva e Ribeiro, 2013 definiu-se altura, comprimento e largura da amostra, indo um pouco mais longe definindo ainda *Dimensões da amostra*. Numa perspetiva de que o investigador optasse por maior especificação, ou por sua vez se o termo mais geral *Dimensões da amostra* o satisfaria. Não foi possível debater esta questão diretamente com um investigador deste domínio pelo que a resposta fica em aberto.

Os autores definiram o descritor *Instrumento* de forma a captar informação contextual sobre os instrumentos utilizados no desenvolvimento das experiências. Por sua vez, da análise foi possível definir instrumentos específicos e características dos mesmos, verificando-se com frequência a referência do nome do instrumento, do produtor entre outras características. Infere-se que é preponderante recolher informação sobre os instrumentos sendo posta em causa, novamente, a questão de uma maior ou menor especificidade.

Relativamente a Método, neste estudo foi possível identificar um método específico *Método de preparação da amostra*. Porém, compreende-se que a especificidade em demasia pode levar a uma perda de informação contextual relevante.

A *Velocidade do teste* é definida com o objetivo de captar valores sobre a velocidade a que o instrumento pressiona o espécimen, para este não foi identificado nenhum descritor que recolha a mesma informação. No entanto, infere-se que poderá ser necessário registar quais os testes aplicados definindo-se para tal descritores como *Teste de avaliação da resistência à fratura* e *Teste de tensão de deformação*.

Os autores definiram que o *Comprimento da fissura antes da experiência* era importante, por sua vez também foi possível constatar que o *Comprimento da fissura* deveria ser anotado. Contudo, não se identificou o momento em que deveria ser capturado esse valor, dando opção ao investigador de anotar o que considerar, sendo que o registo mais ou menos completo será dependente da competência do investigador para a gestão de dados. No caso de a sensibilidade para este assunto ser menor a especificação dos descritores pode ser uma solução.

Por fim, os autores consideraram que as condições de *Temperatura* e *Humidade* eram características relevantes para a contextualização da experiência. Efetivamente, através da análise contextual verificou-se que os investigadores registavam com frequência essas condições, por essa razão o descritor que se propõem é *Condições atmosféricas*, este descritor engloba tanto a temperatura, como a humidade.

## **4.5 Engenharia Mecânica - Processo de fabrico: Moldagem de chapa**

Madeira (2014) acredita que muitos dos avanços tecnológicos impulsionaram o desenvolvimento de novos materiais e inovações nas ligas possibilitando variadas aplicações como é o caso do ramo automóvel. O desenvolvimento da investigação de novos materiais potencializa a evolução e consequentemente a aplicação de materiais com funções diferentes daquelas para as quais foram desenvolvidos inicialmente. São exemplos a aplicação de estruturas metálicas para a construção de edifícios substituindo em parte as soluções tradicionais, como o betão armado; a aplicação em caixilharias, janelas, portas como substituto de madeiras; ou ainda a aplicação em revestimentos de fachadas e coberturas de edifícios. Muitos são os exemplos que se poderiam referir, a tendência é que as novas soluções apresentadas pelos materiais otimizem o tempo de duração, sejam mais económicas, sendo que a segurança deve ser mantida ou aumentada. Para isso é necessário que sejam desenvolvidos os conhecimentos relacionado com as características dos materiais.

Segundo Banabic et al. citado por Ke et al. (2018) a enformabilidade da chapa metálica é a capacidade de sofrer deformação plástica sem que ocorram defeitos.

### **Análise de conteúdo e resultados**

A análise do domínio teve por base um conjunto de artigos que descrevem experiências que estudam as características dos materiais relativamente à sua capacidade de resistência à fratura. A tabela resultante deste estudo pode ser consultada no Anexo IX.

Da interação com o investigador não foi possível avaliar o conjunto de descritores apresentados, uma vez que o investigador achou que uma reflexão sobre aspetos gerais seria um contributo mais pertinente. Iniciando assim a sua primeira reflexão sobre a questão da linguagem, afirmando com alguma persistência que os descritores deveriam ser apresentados aos investigadores em inglês, o que na sua opinião facilitaria a interpretação dos conceitos.

Considerou também que a ordem de apresentação dos descritores ao investigador também é um fator que influencia o entendimento e como tal os descritores devem aparecer organizados por relevância.

Durante a sessão de avaliação o investigador partilhou uma história sobre uma experiência que foi reproduzida em três locais distintos, relatando que as condições foram mais ao menos semelhantes, apesar do instrumento ser apenas diferenciado pelo produtor. Defendeu que os resultados deveriam ter sido os mesmos, no entanto não foi isso que aconteceu, os resultados variaram e não compreendia qual a razão. Perante este relato importa refletir que alguma variável não controlada teve influência na reprodução da experiência, podendo estar relacionada com as características do instrumento. Quando se questionou o investigador sobre o facto do produtor do instrumento ter influência nos resultados, defende que esse fator não deveria interferir. Com este relato enfatiza-se a necessidade de descrever com pormenor as condições experimentais em que são desenvolvidas as experiências, para se necessário reproduzi-las e alcançar os mesmos valores.

Depois das reflexões do investigador foi-lhe pedido a avaliação dos descritores propostos para o contexto experimental que realiza habitualmente. O investigador afirma que no geral os termos estão bem definidos, mas que dizem respeito ao domínio da engenharia mecânica em específico a caracterização dos materiais e que não se ajustam às experiências que desenvolve no domínio da engenharia mecânica, mais focadas nos processos de fabrico. Esta pode ser um potencial desafio na tarefa de selecionar artigos, uma vez que, sem o conhecimento específico a identificação do domínio pode, tal como se verificou, falhar.

Este caso de uso é relevante, primeiro porque reflete a dificuldade do curador em identificar os domínios. Segundo, porque representa o input que é possível efetuar pelo curador de dados, sem a intervenção do investigador. Reafirmando a ideia de que efetivamente é possível

o curador de dados desenvolver um modelo de metadados de forma autónoma, no entanto a avaliação é uma fase preponderante, uma vez que enquanto curador sem conhecimento específico sobre um domínio é fácil escaparem pormenores relevantes.

Durante a entrevista o investigador afirmou que registava o contexto experimental das suas investigações num documento e que podia partilhar, caso se mostrasse relevante para o trabalho. Assim, da análise do documento que relata a experiência de moldagem de chapa, cedido pelo investigador, é possível verificar que é registado todo o contexto experimental, à semelhança do que foi possível verificar no PowerPoint do investigador da física. Nos documentos são registadas imagens, sequências de tarefas, tabelas, todos os procedimentos e aspetos que possam influenciar os dados.

Da análise do documento que contextualiza as experiências de moldagem de chapa foi possível identificar 21 descritores, conclui-se que os requisitos necessários para descrever o contexto experimental são:

TABELA 9-REQUISITOS NECESSÁRIOS PARA A DESCRIÇÃO DE ENGENHARIA MECÂNICA – PROCESSO DE FABRICO:  
MOLDAGEM DE CHAPA

<b>Categoria</b>	<b>Descritores</b>
Propriedades do material	Material/ Código do material/ Dimensões do material/ Data de produção/ Lubrificante do material/ Produtor/ Limpeza do material
Instrumento, características e calibrações	Instrumento/ Histórico de uso da ferramenta/ Desgaste do instrumento/ Rugosidade da superfície do instrumento/ Limpeza dos instrumentos
Método	Método de lubrificação
Outros	Condições ambientais/Teste/ Data do teste /Sistema de aquisição dos dados de teste/ Velocidade de formação/ Força de perfuração/ Deslocação do perfurador/ Força Blankholder

De modo a representar o contexto da experiência é necessário que o investigador registe, qual o *Material* utilizado para o desenvolvimento da investigação, importa registar o código, as dimensões, a data em que foi produzido, qual o parceiro instituição que produziu o material, qual o lubrificante utilizado, como é efetuada a limpeza do material e o que é utilizado para o efeito.

É fundamental registar quais os instrumentos, frequentemente designados por ferramentas, utilizados, qual o histórico de utilização da ferramenta (sempre que a ferramenta tenha sido utilizada em testes anteriores), o desgaste que apresenta, qual a rugosidade da superfície e o tipo de limpeza efetuada para limpar os instrumentos.

No documento estava definido por tópicos o método utilizado para a lubrificação assim deduz-se que é um aspeto importante a registar neste domínio.

É referido ainda que durante os testes experimentais devem ser registadas um conjunto de variáveis que influenciam os resultados. Assim importa descrever qual o teste aplicado, a data em que o mesmo foi realizado, o sistema utilizado para adquirir os dados de teste, as condições ambientais em que se desenvolve a investigação, a velocidade de formação, a força de perfuração, deslocação do perfurador e a força blankholder.

## 5 Discussão

Partindo do princípio de que a análise de conteúdo apresenta características que possibilitam otimizar o trabalho desenvolvido pelo curador de dados, foram analisados cinco casos de uso. Pretendia-se compreender se o curador de dados sem conhecimento prévio em domínios específicos poderia, de forma autónoma, identificar conceitos fundamentais à descrição de dados, que serviriam como ponto de partida para a interação com o investigador. Através da avaliação dos casos de uso junto dos investigadores obteve-se resultados positivos que sustentam esta hipótese.

A taxa de aceitação dos descritores na avaliação do modelo de metadados evidencia as vantagens da análise de conteúdo antes da primeira interação com o investigador. O curador de dados consegue definir potenciais descritores para o modelo de metadados tendo por base uma atividade realista do seu ponto de vista, uma vez que foi possível efetuar a análise de conteúdo num período de tempo razoável, sem exigir um conhecimento aprofundado do domínio. Acredita-se que com a aprendizagem de descritores identificados em domínios anteriores pode tornar a tarefa ainda mais ágil. Nesse sentido defende-se que a análise de conteúdo é uma abordagem válida para melhorar a proatividade do curador de dados. Os investigadores têm agendas ocupadas e o processo de investigação é prioritário, o que leva a valorizar a análise de conteúdo, que requer menos tempo do investigador.

A análise de conteúdo foi baseada apenas nas secções da introdução e secções que relatam o contexto experimental e metodologias de forma a tornar esta tarefa exequível. As secções escolhidas foram analisadas exaustivamente. Constatou-se que é possível analisar um domínio numa semana, despendendo para a realização da tarefa duas horas diárias. Obtendo-se um conjunto de metadados preparado para apresentar ao investigador para posterior validação. Obviamente que cada caso tem as suas especificidades, podendo existir tanto domínios mais complexos bem como mais simples.

Acredita-se que a riqueza desta metodologia assenta na possibilidade de garantir uma interação rica entre o curador e o investigador. Na medida em que, numa primeira fase, o curador efetua a análise que lhe permite ter algumas noções básicas sobre o domínio em causa, facilitando a comunicação. E na segunda fase de avaliação junto do investigador, este com conhecimento específico sobre o domínio pode contribuir de forma a complementar os requisitos de metadados identificados.

A abordagem da análise de conteúdo permite que o investigador compreenda em que consiste a gestão de dados de investigação através da materialização de exemplos práticos do seu domínio, sendo que o comportamento do investigador se transforma, passando de um comportamento mais passivo para um de maior interesse e curiosidade. Através de uma interação em que a comunicação é eficaz a contribuição dos investigadores é maior.



Tal como se verificou nos casos estudados, alguns dos investigadores fizeram sugestões que no seu entender tornariam o processo mais fácil. O investigador da química sustentável, por exemplo, sugere que o processo do curador poderia ser facilitado se fosse possível aceder à informação que os instrumentos utilizados nas experiências recolhem, uma vez que têm campos predefinidos semelhantes ao modelo de metadados apresentado. Uma opinião partilhada pelo investigador da física e pelo investigador do domínio da moldagem de chapa é que a organização dos descritores tornaria a descrição mais fácil.

No entanto, um curador de dados sem conhecimento específico do domínio não é capaz de efetuar essa identificação e respetiva organização dos metadados. A organização que se tentou adotar reuniu descritores relacionados. Um dos grupos considerou as propriedades da amostra, reunindo descritores que de alguma forma se relacionam com a amostra. O outro grupo considerou os descritores que identificam os Instrumentos, as suas características ou calibrações. Criou-se ainda um grupo referente aos tempos registados na realização de alguma atividade, um grupo relativo a temperaturas registadas, um grupo para métodos, outro para técnicas e um grupo designado por outros, onde se agrupavam os restantes descritores para os quais não se identifica uma relação clara com algum dos anteriores. A alocação dos elementos aos grupos de forma a tornar mais clara a apresentação dos resultados não foi sempre evidente. Considerando uma organização mais eficiente, esta deveria ser efetuada por um investigador com conhecimento do domínio, uma vez que tem o conhecimento necessário para compreender as relações entre os conceitos.

Ao longo das análises efetuadas persistiu a dúvida sobre qual o nível de especificidade aceitável, de que forma o modelo de metadados deve ser apresentado ao investigador: com descritores mais genéricos ou por outro lado mais específicos? A definição de um descritor genérico *Instrumento*, ou a especificação dos instrumentos como *Instrumento de análise dos catalisadores*, *Instrumento de medição da radiação eletromagnética* ou a seleção de *Método* em detrimento da especificação dos métodos, *Método de análise*, *Método de síntese da amostra*. A decisão não foi óbvia, por vezes através da análise de conteúdo e com a falta de conhecimento sobre o domínio não foi possível efetuar a especificação; além disso, era evidente que quanto mais se especificasse maior seria a complexidade do modelo de metadados.

Na tentativa de encontrar uma solução foram apresentados ao investigador alguns casos em que estavam previstos descritores genéricos, noutros a especificação e uma situação em que descritores genéricos e específicos apareciam em simultâneo. No momento que o investigador da química se deparou com o descritor *Instrumento* referiu que anotaria o nome de um instrumento, mas que a especificação da função do instrumento tornaria a descrição um processo mais fácil. Por outro lado, quando confrontado com *Método de análise de polifenóis* sugeriu que *Método de análise* seria melhor uma vez que não restringe tanto. No caso do

investigador da energia renovável anotou os descritores *Instrumento* e *Método* sem fazer qualquer tipo de referência a uma necessidade de maior especificidade. Para o caso da física foi possível especificar 20 instrumentos, o investigador referiu a preferência por registar vários tipos de instrumentos quando questionado se teria preferência apenas pelo descritor *Instrumento*. Por sua vez, o investigador da moldagem de chapa sugere ao longo da sessão de avaliação que a generalização é preferida. Isto leva a querer que não existe uma resposta consensual entre os investigadores, sendo necessário antever diferentes cenários.

Por um lado, importa garantir que o investigador não se esquece de descrever toda a informação relevante, orientando para isso os descritores com um maior grau de especificidade, e verificando-se que os investigadores tendem a usá-los com alguma desenvoltura. Por outro lado, deve ser considerado que um modelo de metadados demasiado extenso pode tornar-se impraticável. Na maioria dos casos os investigadores não têm predisposição para descrever demasiados elementos, sendo relatado por um dos investigadores que caso fosse necessário fazer *scroll* (percorrer uma lista) demasiado extensa para descrever os seus metadados isso seria desmotivante. Outra limitação dos modelos de metadados com grande especificidade relaciona-se com a dificuldade na reutilização dos descritores para outros domínios, uma vez que as propriedades capturadas podem mudar significativamente entre subdomínios, o que implica a definição de descritores para cada caso.

De modo a que seja possível a reutilização de descritores, um nível mais geral é preferencial. No entanto, um cenário a evitar é a necessidade dos investigadores por determinado descritor que não foi previsto, ou a situação em que apenas têm à disposição descritores genéricos que podem impedir o registo de informação relevante.

Verifica-se que o desenvolvimento de experiências, que dão origem a dados do tipo experimental, englobam vários métodos, técnicas, ou ainda instrumentos, calibrações, tipos de amostras entre outros. Cada escolha do investigador de determinado instrumento, método ou outro irá influenciar os elementos necessários para efetuar a descrição do conjunto de dados. Exemplificando, o descritor *Distância interfacial* só será necessário se para a amostra escolhida for possível identificar esse valor; *Tempo de secagem da amostra* só é relevante se a amostra sofrer um processo de secagem, e ainda *Potencial de oxidação* é importante se na experiência for utilizado um agente oxidante. O curador de dados deve assim considerar as possíveis variáveis que possam ter influência na configuração experimental.

É necessário desenvolver mais estudos em domínios diversos de forma a compreender possíveis similaridades que poderão possibilitar a partilha de metadados. Uma das hipóteses que se levantou na revisão de literatura é a possibilidade de os requisitos necessários para a descrição de conjuntos de dados terem características diferentes que se relacionam com cada tipologia de dados. Para os casos analisados em que os dados são experimentais verifica-se que há requisitos

comuns como a necessidade de identificar a amostra, espécimen ou material em estudo, as suas propriedades, definir quais os instrumentos, características e calibrações utilizados nas experiências, registar informação relacionada com tempos e temperaturas e ainda métodos e técnicas. Comprova-se que existe similaridade entre os domínios analisados, ainda assim é necessário colecionar mais casos.

Outra das reflexões é a necessidade de encontrar um equilíbrio no desenvolvimento de modelos de metadados, uma vez que a facilidade em definir com alguma rapidez os modelos, tendo por base a análise de conteúdo, pode levar a uma proliferação de modelos que se pode tornar caótica. Acredita-se que a reutilização de descritores entre domínios pode ser dificultada pelo excesso de modelos, por sua vez, a partilha de dados poderá ser facilitada pela reutilização de descritores.

A análise de conteúdo pode ser generalizada e aplicada em outras atividades no contexto dos metadados como o desenvolvimento de ontologias.

Ainda a aplicação da abordagem através do processo automático. Entende-se que abordagem manual e automática podem na fase exploratória, desenvolver-se em paralelo e reforçar a colaboração. A recolha automática de conceitos acelera o processo de análise, contudo, para o entendimento dos conceitos extraídos automaticamente é necessário que o curador de dados tenha alguma experiência do domínio.

A análise de conteúdo manual pode ser aplicada sempre que for necessário definir um modelo de metadados de raiz, ou para especializar ferramentas existentes.

## 6 Conclusões

Através das avaliações efetuadas verificar-se que a análise de conteúdo apresenta características que otimizam o trabalho do curador de dados. Com base na análise dos vários domínios foi possível compreender que existem conceitos que se repetem, o que possibilita a reutilização de descritores; a cada caso analisado aumenta em parte a habilidade do curador em identificar conceitos, reduzindo assim o esforço na análise de novos domínios.

Considera-se que a abordagem utilizada para identificar conceitos dos domínios que apoiam o desenvolvimento de modelos de metadados tem vários aspetos positivos, mas ainda há espaço para melhorias. Será necessário desenvolver mais estudos para precaver ou solucionar os aspetos negativos. A Tabela 10 sumariza os aspetos positivos e negativos da abordagem.

TABELA 10-RESUMO DOS ASPETOS POSITIVOS E NEGATIVOS DA ANÁLISE DE CONTEÚDO

Aspetos positivos	Aspetos negativos
Trabalho proativo do curador de dados; Melhoria na comunicação entre o curador e o investigador; Facilidade na definição de um modelo de metadados ajustado às necessidades dos domínios num período de tempo reduzido.	Dificuldade em definir domínios; Dificuldade em definir alguns descritores pela falta de conhecimento específico sobre o domínio; Dificuldade em equilibrar a especificidade com a generalidade.

Esta abordagem pode ser vista como um input inicial em que o curador sem conhecimento sobre um domínio efetua a análise de conteúdo. Assim, o curador pode compreender quais os conceitos fundamentais para a sua descrição para que na primeira interação com o investigador a comunicação seja facilitada. O curador está preparado para apresentar exemplos práticos de metadados no domínio do investigador, melhorando a perceção do investigador sobre a gestão de dados de investigação, o que o leva a contribuir ativamente com sugestões ou correções. Entende-se a análise de conteúdo como uma preparação para a interação com o investigador, compreendendo-se que é uma abordagem que possibilita a definição de modelos de metadados, num curto período de tempo. A tendência é melhorar a abordagem de modo a torna-la menos exhaustiva. Os resultados obtidos no caso da física mostram que é possível escolher poucos descritores, alguns mais genéricos também, e deixar que o investigador proponha as suas necessidades.

Contudo são identificadas algumas limitações no processo. Uma vez que não se tem conhecimento específico, a delimitação dos domínios torna-se uma tarefa difícil. Outra dificuldade é o entendimento de determinado excerto de texto, por exemplo, quando no artigo são referidos conceitos como polifenol ou carbono ativado, o curador sem o conhecimento específico do domínio sente dificuldade em identificar se os autores se referem a poluentes, a

catalisadores ou outras categorias de substâncias. Neste sentido é essencial o envolvimento do investigador na validação e desenvolvimento de modelos de metadados. Porém esta avaliação depende da disponibilidade e sensibilidade do investigador para a gestão de dados de investigação.

A dificuldade passa por compreender até que ponto é benéfica a especificação. Quando o conjunto de descritores é demasiado extenso e específico pode 1) limitar a participação do investigador; 2) tornar-se impraticável. Por outro lado, quando demasiado genérico pode transformar-se numa tarefa exaustiva para o investigador. Na medida em que tem de refletir sobre quais os conceitos necessários a descrever.

## **6.1 Limitações do estudo**

O tempo para o estudo foi limitado ao período do projeto de dissertação de mestrado que durou aproximadamente seis meses, pelo que apenas foi possível a análise de cinco casos de uso. A coleção de mais casos seria importante para compreender melhor a questão da reutilização dos descritores entre domínios.

O tempo foi um fator que também influenciou as avaliações, sendo que o conjunto de descritores escolhidos foram avaliados apenas por um investigador, por domínio. Dada a impossibilidade de controlar a sensibilidade e interesse dos investigadores para esta avaliação, podemos ter vários cenários 1) dificuldade para verificar que conceitos se repetem no modelo de metadados; 2) descritores importantes por preencher; 3) registo de descritores não relevantes; 4) rejeição de descritores que o investigador não conhece e não utiliza. Seria, portanto, interessante efetuar mais avaliações junto de investigadores do mesmo domínio.

A falta de conhecimento específico sobre os domínios é outra das condicionantes do curador de dados, implicando uma possível deficiência na definição dos domínios como se verificou no caso da moldagem de chapa. Assim reafirma-se a importância de integrar os investigadores no processo de gestão de dados, como *stakeholders* principais, e de desenvolver o modelo de metadados de modo a que o investigador o possa usar. À semelhança do desenvolvimento de um produto para um cliente também o modelo de metadados deve ser entendido como um produto que deve satisfazer as necessidades dos investigadores e para isso deve ser criado a pensar na facilidade de utilização. Assim é fundamental compreender qual o grau de especificidade que agrada aos investigadores e permite recolher toda a informação necessária à descrição do contexto experimental.

## **6.2 Investigação futura**

De modo a complementar a investigação poderia ser relevante efetuar um mapeamento sobre modelos de metadados existentes de forma a compreender se seria possível reutilizar alguns descritores anteriormente definidos. Outra linha de trabalho futuro pode basear-se no desenvolvimento de ontologias a partir dos modelos de metadados definidos.

A coleção de mais casos de uso avaliados junto de vários investigadores permitiria compreender com maior segurança se os resultados obtidos são válidos e se a generalização dos requisitos pode abranger as várias tipologias de dados. Esta extensão também pode contribuir para encontrar consenso para o problema da maior ou menor especificação dos descritores.

A relevância e o interesse pela abordagem utilizada neste trabalho deram origem à aceitação de um artigo, que relata o estudo desenvolvido que será publicado no 13º Congresso BAD. A abordagem, comentada pelo revisor como reflexão de interesse, motiva o desenvolvimento de mais estudos nesta área, como exemplo investigações que relacionem o processo automatizado com a análise de conteúdo manual.

## 7 Referências Bibliográficas

- Amorim, Ricardo Carvalho, João Aguiar Castro, João Rocha da Silva, e Cristina Ribeiro. 2016. «A comparison of research data management platforms: architecture, flexible metadata and interoperability». *Universal Access in the Information Society*. <https://doi.org/10.1007/s10209-016-0475-y>.
- Assante, Massimiliano, Leonardo Candela, Donatella Castelli, e Alice Tani. 2016. «Are Scientific Data Repositories Coping with Research Data Publishing?» *Data Science Journal* 15 (6):1–24. <https://doi.org/10.5334/dsj-2016-006>.
- Awre, Chris, Jim Baxter, Brian Clifford, Janette Colclough, Andrew Cox, Nick Dods, Poul Drummond, et al. 2015. «Research Data Management as a “wicked problem”». *Library Review*, 64 (4/5):356–71. <https://doi.org/10.1108/LR-04-2015-0043>.
- Borgman, Christine. 2012. «Advances in information Science: The Conundrum of Sharing Research Data». *Journal of the American Society for Information Science and Technology*, 1059–1078. <https://doi.org/10.1002/asi>.
- Borgman Christine, Wallis Jillian, Mayernik Matthew, Pepe Alberto. 2007. «Drowning in Data: Digital Library Architecture to Support Scientific Use of Embedded Sensor Networks». *JCDL*, 2007. <https://doi.org/10.1145/1255175.1255228>.
- Castro, João. 2013. «Estudo de utilização do Repositório de dados da Universidade do Porto». Universidade do Porto. <http://hdl.handle.net/10216/68647>.
- Castro, João Aguiar, Deborah Perrotta, Ricardo Carvalho Amorim, João Rocha da Silva, e Cristina Ribeiro. 2015. «Ontologies for Research Data Description: A Design Process Applied to Vehicle Simulation», 348–54. [https://doi.org/10.1007/978-3-319-24129-6\\_30](https://doi.org/10.1007/978-3-319-24129-6_30).
- Castro, João Aguiar, João Rocha da Silva, e Cristina Ribeiro. 2013. «Designing an Application Profile Using Qualified Dublin Core: A Case Study with Fracture Mechanics Datasets». *Proc. of the International Conference on Dublin Core and Metadata Applications*, 47–52. <http://dcevents.dublincore.org/IntConf/dc-2013/paper/view/180>.
- Chao, Tiffany C. 2014. «Identifying Indicators of Description for Research Data from Scientific Journal Publications». *iConference 2014 Proceedings*. <http://doi.org/10.9776/14366>.
- Cisar, Petr, Dmytro Soloviov, Antonin Barta, Jan Urban, e Dalibor Stys. 2016. «BioWes-from design of experiment, through protocol to repository, control, standardization and back-tracking». *BioMedical Engineering Online* 15 (1). BioMed Central:129–47. <https://doi.org/10.1186/s12938-016-0188-8>.
- Dahak, Mustapha, Nouredine Touat, e Nouredine Benseddiq. 2017. «On the classification of normalized natural frequencies for damage detection in cantilever beam». *Journal of Sound and Vibration* 402. Elsevier Ltd:70–84. <https://doi.org/10.1016/j.jsv.2017.05.007>.
- Day, Michael. 2005. «DCC Digital Curation Manual Instalment on Metadata». *HATII*, University of

- Glasgow; University of Edinburgh; UKOLN, University of Bath; Council for the Central Laboratory of the the Research Councils.  
<http://www.era.lib.ed.ac.uk/handle/1842/3321%5Cnhttp://www.dcc.ac.uk/resource/curation-manual/chapters/metadata>.
- European Commission. 2017. «H2020 Programme: Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020».
- European Research Council. 2017. «European Research Council (ERC) Guidelines on Implementation of Open Access to Scientific Publications and Research Data in projects supported by the European Research Council under Horizon 2020».
- Ke, Junyi, Yuqi Liu, Hongchuan Zhu, e Zhibing Zhang. 2018. «Formability of sheet metal flowing through drawbead—an experimental investigation». *Journal of Materials Processing Technology* 254 (November 2017). Elsevier:283–93. <https://doi.org/10.1016/j.jmatprotec.2017.11.051>.
- Lyon, Liz. 2007. «Dealing with Data: Roles, Rights, Responsibilities and Relationships Consultancy Report». *Archives*, n. June:1–65.
- Madeira, Tânia Ferreira. 2014. «Caracterização de chapa metálica para a Indústria Automóvel». Instituto Superior Técnico. [https://fenix.tecnico.ulisboa.pt/downloadFile/1970719973965957/Dissertacao\\_Tania\\_Madeira.pdf](https://fenix.tecnico.ulisboa.pt/downloadFile/1970719973965957/Dissertacao_Tania_Madeira.pdf).
- Møberg, Alex, Michal K. Budzik, e Henrik M. Jensen. 2017. «Crack front morphology near the free edges in double and single cantilever beam fracture experiments». *Engineering Fracture Mechanics* 175. Elsevier Ltd:219–34. <https://doi.org/10.1016/j.engfracmech.2017.01.030>.
- NSF. 2005. «Long-lived Digital Data Collections: Enabling Research and Education in the 21st Century». [http://www.nsf.gov/nsb/documents/2005/LLDDC\\_report.pdf](http://www.nsf.gov/nsb/documents/2005/LLDDC_report.pdf).
- OECD. 2007. «OECD Principles and Guidelines for Access to Research Data from Public Funding». França. [http://www.nsf.gov/nsb/documents/2005/LLDDC\\_report.pdf](http://www.nsf.gov/nsb/documents/2005/LLDDC_report.pdf).
- . 2015. *Frascati Manual 2015: Guidelines for collecting and reporting data on research and experimental development*. Paris: OECD. <https://doi.org/10.1787/9789264239012-en>.
- Heidorn, P Bryan. 2008. «Shedding Light on the Dark Data in the Long Tail of Science». *Library Trends* 57 (2):280–99. <https://doi.org/10.1353/lib.0.0036>.
- Qin, Jian, Alex Ball, e Jane Greenberg. 2012. «Functional and architectural requirements for metadata: supporting discovery and management of scientific data». *Twelfth International Conference on Dublin Core and Metadata Applications*, 62–71. <https://doi.org/10.1007/s00799-013-0106-7>.
- RCAAP. 2010. «Os Repositórios De Dados Científicos: Estado da arte».
- Ribeiro, Cristina, João Rocha, Da Silva, João Aguiar Castro, Ricardo Carvalho Amorim, e João Correia Lopes. 2016. «TAIL». *Cadernos BAD*. n. 2:256–64.



- Riley, Jenn. 2017. *Understanding Metadata: What Is Metadata, and What is it for? NISO Primer*. National Information Standards Organization (NISO). <https://doi.org/10.1017/S0003055403000534>.
- Sargent, Robert G. 2011. «Verification and validation of simulation models». Em *Proceedings of the 2011 Winter Simulation Conference*, 183–98. Syracuse University. <https://doi.org/10.1109/WSC.2011.6148117>.
- Smit, Eefke, Jeffrey Van der Hoeven, e David Giarretta. 2011. «Avoiding a digital dark age for data: Why publishers should care about digital preservation». *Learned Publishing* 24 (1):35–49. <https://doi.org/10.1087/20110107>.
- Thanos, Costantino. 2017. «Research Data Reusability: Conceptual Foundations, Barriers and Enabling Technologies». *Publications* 5 (1):19. <https://doi.org/10.3390/publications5010002>.
- UK Data Archive. 2011. «Managing and Sharing Data Best Practice for Researchers».
- University of Essex, University of Manchester and Jisc. 2017a. «UK Data Service Catalogue metadata». 2017. <https://www.ukdataservice.ac.uk/manage-data/document/metadata>.
- . 2017b. «UK Data Service Document your data». 2017. <https://www.ukdataservice.ac.uk/manage-data/document>.
- Valverde-Alva, M. A., T. García-Fernández, M. Villagrán-Muniz, C. Sánchez-Aké, R. Castañeda-Guzmán, E. Esparza-Alegría, C. F. Sánchez-Valdés, J. L. Sánchez Llamazares, e C. E. Márquez Herrera. 2015. «Synthesis of silver nanoparticles by laser ablation in ethanol: A pulsed photoacoustic study». *Applied Surface Science* 355:341–49. <https://doi.org/10.1016/j.apsusc.2015.07.133>.
- Wallis, Jillian C., Elizabeth Rolando, e Christine L. Borgman. 2013. «If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology». *PLoS ONE* 8 (7). <https://doi.org/10.1371/journal.pone.0067332>.
- White, Hollie C. 2014. «Descriptive Metadata for Scientific Data Repositories: A Comparison of Information Scientist and Scientist Organizing Behaviors». *Journal of Library Metadata* 14 (1):24–51. <https://doi.org/10.1080/19386389.2014.891896>.
- Willis Craig, Greenberg Jane, White Hollie. 2012. «Analysis and Synthesis of Metadata Goals for Scientific Data». *Journal of the American Society for Information Science and Technology*, 1505–1520. <https://doi.org/DOI: 10.1002/asi>.
- Wilson, Amanda. 2007. «Toward Releasing the Metadata Bottleneck: A Baseline Evaluation of Contributor-supplied Metadata». *Association for Library Collections & Technical Services* 51. <https://journals.ala.org/index.php/lrts/article/view/5384/6604>.

## Anexo I. Tabela de análise do Caso de Uso Química sustentável - Remoção de partículas poluentes

Expressão em linguagem natural que suscita interesse	Sugestão de descritores em linguagem natural	Nome da seção analisada (atribuído pelo autor)
Lucas, M. S., Peres, J. A., & Li Puma, G. (2010). Treatment of winery wastewater by ozone-based advanced oxidation processes (O3, O3/UV and O3/UV/H2O2) in a pilot-scale bubble column reactor and process economics. Separation and Purification Technology, 72(3), 235–241. <a href="https://doi.org/10.1016/j.seppur.2010.01.016">https://doi.org/10.1016/j.seppur.2010.01.016</a>		
“standard oxidation potential of the hydroxyl radical ( $E_0 = 2.80$ V) is much higher than that of ozone ( $E_0 = 2.07$ V)”	Potencial de oxidação do agente oxidante	Introduction
“interfacial area, $m^2/m^3$ ”	Área interfacial	
“chemical oxygen demand, $g\ L^{-1}$ ”	Carência química de oxigênio	
“individual volumetric mass transfer coefficient, $ms^{-1}$ ”	Coeficiente de transferência de massa volumétrica individual	
“ozone mass flow rate, $g\ s^{-1}$ ”	Taxa de fluxo de massa de ozono	
“ozone interfacial concentration, $g\ L^{-1}$ ”	Concentração interfacial de ozono	
“ozone partial pressure in gas phase, kPa”	Pressão parcial de ozono na fase gasosa	
“volumetric flow rate of gas phase, $L\ s^{-1}$ ”	Taxa de fluxo volumétrico na fase gasosa	
“total organic carbon, $g\ L^{-1}$ ”	Carbono orgânico total	
“ozonation time, s”	Tempo de ozonização	
“superficial gas velocity, $ms^{-1}$ ”	Velocidade superficial do gás	Materials and methods
“The pH”	pH do poluente	
“the absorbance at 254 nm”	Absorbância	
“the total carbon (TC)”	Carbono total	
“inorganic carbon (IC)”	Carbono inorgânico	
“polyphenols”	Polifenóis	
“The ozonation and the experiments with ozone-based AOPs (O3/UV and O3/UV/H2O2)”	Agente oxidante	

“The ozonation and the experiments with ozone-based AOPs (O3/UV and O3/UV/H2O2) were conducted in a bubble-column, semi-batch reactor (internal diameter 0.1 m, height 1m) made of QVF borosilicate glass equipped with a quartz UV-lamp tube (external diameter 40mm)”	Reator de ozonização	
“it was controlled using using 0.1M NaOH or 0.1M H2SO4 solutions”	Solução de controlo	
“The absorbance of samples was determined by a UV/Vis Spectrophotometer, UVmini 1240 Shimadzu.”	Instrumento de medição da absorbância	
“The Folin-Ciocalteu method, using gallic acid as a standard, was followed to evaluate the total polyphenols content in the samples.”	Método de análise de polifenóis	
Sobana, N., & Swaminathan, M. (2007). Combination effect of ZnO and activated carbon for solar assisted photocatalytic degradation of Direct Blue 53. <i>Solar Energy Materials and Solar Cells</i> , 91(8), 727–734. <a href="https://doi.org/10.1016/j.solmat.2006.12.013">https://doi.org/10.1016/j.solmat.2006.12.013</a>		
“In the present work we report the photocatalytic degradation of DB53 with 9AC–ZnO using solar light”	Técnica de remoção de partículas	Introduction
“the photocatalytic degradation of DB53 with 9AC–ZnO using solar light”	Composto químico	
“ZnO (surface area 5m2 g <sub>-1</sub> , particle size 4.80 mm)”	Área da superfície do catalisador	Experimental
“ZnO (surface area 5m2 g <sub>-1</sub> , particle size 4.80 mm)”	Tamanho da partícula do catalisador	
“Activated carbon from SD fine chemicals”	Adsorvente / Adsorção	
“Activated carbon from SD fine chemicals (surface area 735.60m2 g <sub>-1</sub> )”	Área da superfície do adsorvente	
“Activated carbon from SD fine chemicals (...) ash content 2.5%”	Teor de cinzas do adsorvente	
“Activated carbon from SD fine chemicals (...) particle size 300 mesh”	Tamanho da partícula do adsorvente	
“Activated carbon from SD fine chemicals (...) and impurities content of acid solubles 2.5%, water solubles 1.5%”	Teor de impurezas	
“Molecular formula=C34H24N6Na4O14S4”	Formula molecular do adsorvente	
“molecular weight =960.82”	Massa molecular	
“Catalysts with activated carbon were obtained by mixing ZnO and activated carbon	Reagentes	

at different proportions in an aqueous suspension and this was continuously stirred for 3 h.”		
“Catalysts with activated carbon were obtained by mixing ZnO and activated carbon at different proportions in an aqueous suspension and this was continuously stirred for 3 h.”	Tempo de agitação da suspensão	
“After this, the mixture was filtered and dried at room temperature”	Temperatura de secagem da amostra	
“The characterization of the AC–ZnO catalysts using XRD analysis and surface area measurements has been reported”	Catalisador	
All photocatalytic experiments were carried out under similar conditions on sunny days of March–April 2004 between 11AM and 2 PM.	Condições atmosféricas	
“An open borosilicate glass tube of 50 ml capacity, 40 cm height and 20mm diameter was used as a reaction vessel with the total light exposure length of 330 mm.”	Vaso de reação fotocatalítica	
“BET specific surface área ( $\text{m}^2 \text{g}^{-1}$ )”	Área da superfície da amostra	
“Crystallite size (nm)”	Tamanho da cristalite da amostra	
“Pore specific volume ( $\text{cm}^3 \text{g}^{-1}$ )”	Volume do poro adsorvente	
“UV spectral measurements were done using Hitachi U- 2001 spectrophotometer.”	Instrumento de medição espectral	
“The pH of the solution is measured by using HANNA Phep (Model H 198107) digital pH meter.”	Instrumento de medição do pH	
“Solar light intensity was measured for every 30 min and the average light intensity over the duration of each experiment was calculated.”	Tempo de medição da intensidade da luz	
“The intensity of solar light was measured using LT Lutron LX-10/A digital Lux meter.”	Instrumento de medição da intensidade da luz	
Sobana, N., Muruganandam, M., & Swaminathan, M. (2007). Characterization of AC–ZnO catalyst and its photocatalytic activity on 4-acetylphenol degradation. <i>Catalysis Communications</i> , 9(2), 262–268. <a href="https://doi.org/10.1016/j.catcom.2007.04.040">https://doi.org/10.1016/j.catcom.2007.04.040</a>		
“The semiconductor mediated heterogeneous photocatalysis is an interesting and promising	Técnica de remoção de partículas	Introduction

technique for environmental cleaning and remediation due to its potential to destroy a wide range of organic and inorganic pollutants at ambient temperatures and pressures.”		
“surface area 5 m <sup>2</sup> g <sup>-1</sup> ”	Área da superfície do catalisador	Experimental
“particle size 4.80 $\mu$ m”	Tamanho da partícula do catalisador Tamanho da partícula do poluente	
“surface área 735.60 m <sup>2</sup> g <sup>-1</sup> ”	Área da superfície do adsorvente	
“ash content 2.5%”	Teor de cinzas do adsorvente	
“particle size 300 mesh”	Tamanho da partícula do adsorvente	
“impurities content”	Teor de impurezas	
“were obtained by mixing ZnO with activated carbon at different proportions in an aqueous suspension and continuous stirring for 3 h”	Tempo de agitação da suspensão	
“After this, the mixture was filtered and dried.”	Temperatura de secagem da amostra	
“Scanning electron microscopic (SEM) analysis was performed on platinum coated catalysts using a Jeol apparatus model JSM-5610 LV.”	Instrumento de análise dos catalisadores	
“A Varian Cary 5E UV/VIS–NIR spectrophotometer equipped with an integrated sphere was used to record the diffuse reflectance spectra (DRS) and to measure the absorbance data of the solution samples.”	Instrumento de medição da radiação eletromagnética	
“The baseline correction was performed using a calibrated reference sample of barium sulfate.”	Amostra de referência	
“wavelength range of 200–800 nm”	Comprimento de onda	
“The specific surface area of the catalysts was determined through nitrogen adsorption at 77 K on the basis of BET”	Técnica de medição da área superficial	
“The specific surface area of the catalysts was determined through nitrogen adsorption at 77 K on the basis of BET equation using a sorptomatic 1990 instrument.”	Instrumento para a medição da área superficial	

“Photocatalytic activity of the catalyst was determined by the degradation of 4-acetylphenol in water under UV-irradiation”	Atividade fotocatalítica	
“added into 50 mL of aqueous solution of 4-AP ( $3 \times 10^{-4}$ mol L <sup>-1</sup> )”	Solução aquosa	
The solution was irradiated with UV light using Heber multilamp photoreactor model HML-MP 88 which consists of eight medium pressure mercury vapor lamps (8 W) set in parallel and emitting 365 nm wavelength”	Instrumento de radiação de luz UV	
To monitor the concentration of 4-AP, 2 mL of the sample was withdrawn at specific time intervals and centrifuged to remove the catalyst.	Quantidade da amostra centrifugada	
One millilitre of the centrifugate was diluted to 10 mL and its absorbance at 276 nm was measured.	Quantidade da amostra centrifugada diluída	
“its absorbance at 276 nm was measured”	Absorvência	
“Solar light intensity was measured for every 30 min and the average light intensity over the duration of each experiment was calculated”	Tempo de medição da intensidade da luz	
“The intensity of solar light was measured using LT Lutron LX-10/A digital Lux meter”	Instrumento de medição da intensidade da luz	
The intensity of solar light (1100 · 100 lx) was nearly constant during the experiments.	Intensidade da luz solar	
“UV spectral measurements were done using Hitachi U-2001 spectrophotometer”	Instrumento de medição espectral	

## Anexo II. Tabela de resultados do Caso de Uso Química sustentável- Remoção de partículas poluentes

Input do curador de dados		Input do investigador		
Categorização de descritores	Descritores apresentados ao investigador	Anotação	Comentário	Sugestão de melhoria
Propriedades da amostra	Composto químico		Causa dúvida de interpretação, tudo é um composto químico e assim não tem especificidade.	Amostra
	Elemento químico	Titânio (Ti)		
	Coefficiente de transferência de massa volumétrica individual		A investigadora não sabia o que responder, deveria ser algo mais específico.	Quantidade do composto degradado
	Tamanho da cristalite da amostra	10 nm		
	Volume do poro da amostra	20 cm <sup>3</sup>		
	Área interfacial		Nem todas as amostras tem uma área interfacial	Distância interfacial
	Carência química de oxigênio		Não faz muito sentido desta forma, os dois são	Gás em estudo
	Taxa de fluxo volumétrico na fase gasosa		parecidos. Assim primeiro é necessário saber qual o gás estudado.	Fluxo da fase gasosa
	Velocidade superficial do gás		Não diz nada, é definido pelo (fluxo da fase gasosa)	
	Carbono inorgânico		Não sabe dizer um valor porque não utiliza, mas	

			reconhece o conceito.	
	Polifenóis		<p>“não percebo, queres quantidade ou formula”</p> <p>Para a quantidade deveria ser questionado massa/volume do polifenol, se por outro lado, se a questão é qual o polifenol usado poderia ser formula do polifenol</p>	<p>Massa do polifenol</p> <p>Fórmula do polifenol</p>
	Agente oxidante	Oxigénio		
	Potencial de oxidação do agente oxidante	3.2 eV	Se o descritor fosse potencial de oxidação a resposta seria a mesma, uma vez que o agente oxidante estava definido.	Potencial de oxidação
	Massa molecular		É necessário definir de que é a massa molecular.	<p>Massa molecular do polifenol;</p> <p>Massa molecular do gás;</p> <p>(...)</p>
	Reagentes		Coloca-se tudo que está a ser usado na experiência	
	Teor de impurezas	99%	É muito específico, normalmente é anotado antes o grau de pureza e o teor de impureza é obtido através de cálculos.	Grau de pureza
	Taxa de fluxo de massa de ozono		Repete o conceito do Fluxo da fase gasosa.	



	Concentração interfacial de ozono		Não consegue perceber	
	Pressão parcial de ozono na fase gasosa		É um intervalo de valores, como não lida com esta questão não sabe dar um valor. Com base neste intervalo é possível calcular a área superficial e o volume de poros	
	Carbono orgânico total		Não lida com este composto, mas sabe que admite valores muito elevados.	
	Absorbância	450 nm		
	Carbono total		É a soma do carbono orgânico e do inorgânico	
	pH do poluente	6,5		pH da solução
	Tamanho da partícula do poluente			Tamanho da partícula
	Catalisador	TiO <sub>2</sub>		
	Área da superfície do catalisador	80cm <sup>2</sup> /g		
	Tamanho da partícula do catalisador			Tamanho da partícula
	Comprimento de onda do catalisador		Não faz sentido	
	Adsorvente	Carvão ativado		
	Área da superfície do adsorvente	1000 cm <sup>2</sup> /g		
	Teor de cinzas do adsorvente	2,8%		
	Tamanho da partícula do adsorvente	80nm		
	Formula molecular do adsorvente	Carbono		
	Amostra de referência	TiO <sub>2</sub>		
	Solução de controlo	NaCl		

	Solução aquosa		“aquosa” repete a ideia de solução, por isso é desnecessário. É necessário definir que tipo de solução é.	Solução de limpeza; Solução ácida; (...)
	Absorvência		Não sabe o que é	
<b>Instrumento, características e calibrações</b>	Instrumento		Responderia com um nome de um instrumento, contudo se especificar a função do instrumento é mais fácil.	Especificação do instrumento como aparece nas linhas a baixo.
	Reator de ozonização		Está correto, faz sentido.	
	Instrumento de medição da absorvência		Está correto, faz sentido.	
	Instrumento de radiação de luz UV		Está correto, faz sentido.	
	Instrumento de medição da intensidade da luz		Está correto, faz sentido.	
	Instrumento de medição espectral		Não sabe.	
	Instrumento de medição do pH		Está correto, faz sentido.	
	Instrumento de análise dos catalisadores		Está correto, faz sentido.	
	Instrumento de medição da radiação eletromagnética		Está correto, faz sentido.	
	Instrumento para a medição da área superficial		Está correto, faz sentido.	
	Vaso de reação fotocatalítica		Está correto, faz sentido. É identificado pelo volume que suporta.	

	Quantidade da amostra centrifugada	50ml		
	Quantidade da amostra centrifugada diluída		Repete a ideia da anterior “Quantidade da amostra centrifugada”	
<b>Tempo</b>	Tempo de ozonização		Faz sentido, qualquer valor em horas ou minutos	
	Tempo de medição da intensidade da luz		Faz sentido, qualquer valor em horas ou minutos	
	Tempo de agitação da suspensão		Faz sentido, qualquer valor em horas ou minutos	
<b>Temperatura</b>	Temperatura de secagem da amostra	100°C		
<b>Método</b>	Método de análise de polifenóis	DQO (Demanda química de oxigénio)	Seria melhor colocar só método de análise porque não restringe.	Método de análise
<b>Técnica</b>	Técnica de remoção de partículas	Fotocatálise		
	Técnica de medição da área superficial	Isotérmicas de adsorção e dessorção de azoto		
<b>Outros</b>	Condições atmosféricas	Condições ambiente	Em outros casos a resposta pode variar para a pressão e a temperatura	
	Atividade fotocatalítica	80%		
	Intensidade da luz solar		Faz sentido, mas não se lembra de um valor	

### Anexo III. Tabela de análise do Caso de Uso Energias Renováveis - Produção de camadas absorventes: Filmes finos

Expressão em linguagem natural que suscita interesse	Sugestão de descritores em linguagem natural	Nome da secção analisada (atribuído pelo autor)
“Investigation of the optical properties of CuInSe <sub>2</sub> and CuInS <sub>2</sub> thin films for photovoltaic application”		
“band gap which can hold values between 1.018 and 1.701 eV”	Intervalo da banda	Introduction
“a high absorption coefficient in visible light and near infrared of about 10 <sup>5</sup> cm <sup>-1</sup> for 1.5 eV”	Coeficiente de absorção	
“yield up to (20.3%)”	Rendimento do composto	
“There are several processes for making the absorbent layers, electroplating is one among them”	Técnica de produção de camadas absorventes	
“CIS (CuInSe <sub>2</sub> and CuInS <sub>2</sub> )”	Elemento químico	Experimental
“CIS (CuInSe <sub>2</sub> and CuInS <sub>2</sub> )”	Composto químico	
“elements often in the form (CuCl <sub>2</sub> , InCl <sub>3</sub> , H <sub>2</sub> SeO <sub>3</sub> )”	Estado físico do composto	
“In our case the deposition potentials of the Cu, In, and Se or S elements are diferente”	Potencial de deposição	
“using 0.1 M of the sodium citrate which was chosen to be a complexing agente”	Agente complexante	
“chosen to be a complexing agent in deionized (DI) water”	Solvente	
“the electrochemical deposition reaction”	Tipo de reação	
“the electrochemical deposition reaction of the elements is given by Eq.”	Equação química	
“Conductive indium tin oxide (ITO) coated glass substrates of 1.5 × 1 cm <sup>2</sup> dimension were used as substrates.”	Tipo de Substrato	
“Conductive indium tin oxide (ITO) coated glass substrates of 1.5 × 1 cm <sup>2</sup> dimension were used as substrates.”	Dimensão do substrato	
“They were ultrasonically cleaned with ethanol and deionized water”	Tipo de Limpeza do substrato	
“They were ultrasonically cleaned (...) during 10 min”	Tempo de limpeza	

“The electrodeposition technique was carried out potentiostatically using an Autolab potentiostat/galvanostat”	Instrumento	
“The used working electrode was ITO-coated glass substrate”	Eléctrodo de trabalho	
“the reference electrode was an Ag/AgCl (3 M NaCl)”	Eléctrodo de referência	
“and a platinum plate was used as a counter electrode.”	Contra eléctrodo	
“The composition of the deposition bath consisted of 3 mM of $\text{InCl}_2$ , 3 mM of $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$ , 6 mM of $\text{HSeO}_2$ and 0.1 M of the sodium citrate”	Composição do banho	
“The cathodic potential and deposition time have been fixed at $-750$ mV and 60 min”	Potencial catódico	
“The cathodic potential and deposition time have been fixed at $-750$ mV and 60 min”	Tempo de deposição	
“dried in $60^\circ\text{C}$ oven for 5 min.”	Temperatura de secagem da amostra	
“dried in $60^\circ\text{C}$ oven for 5 min.”	Tempo de secagem da amostra	
“Due to improvement crystallinity of the deposited films and their amorphous nature, all as-deposited films have been annealed in tubular annealing at $350^\circ\text{C}$ ”	Temperatura de recozimento	
“Due to improvement crystallinity of the deposited films and their amorphous nature, all as-deposited films have been annealed in tubular annealing at $350^\circ\text{C}$ during 5 min”	Tempo de recozimento	
“It was necessary to stabilize the temperature for 10 min before to put the samples.”	Tempo de estabilização da temperatura	
“Various characterization techniques, such as optical microscopy, were used”	Técnica de caracterização	
“UV-Visible for extracting optical parameters to know gap energy”	Energia do intervalo	
“UV-Visible for extracting optical parameters to know (...) refractive index”	Índice de refração	
UV-Visible for extracting optical parameters to know (...) extinction coefficient	Coefficiente de extinção	

UV-Visible for extracting optical parameters to know (...) dielectric constant.	Constante dielétrica	
Seol, J.-S., Lee, S.-Y., Lee, J.-C., Nam, H.-D., & Kim, K.-H. (2003). Electrical and optical properties of Cu <sub>2</sub> ZnSnS <sub>4</sub> thin films prepared by rf magnetron sputtering process. <i>Solar Energy Materials and Solar Cells</i> , 75(1), 115–162. <a href="https://doi.org/10.1016/S0927-0248(02)00127-7">https://doi.org/10.1016/S0927-0248(02)00127-7</a>		
“CuInS <sub>2</sub> (CIS) is the chalcopyrite-type semiconductors”	Tipo de semiconductor	Introduction
“has a direct energy band gap (E <sub>g</sub> ¼ 1:08 eV)”	Intervalo da banda	
“a large optical absorption coeficiente”	Coeficiente de absorção	
lack of the detailed understanding of the parameters governing the performance and low-efficiency	Rendimento do composto	
“CZTS thin films were deposited on corning 7059 glass substrates”	Substrato	Experimental
“(…) without substrates heating”	Temperatura do substrato	
“Cu <sub>2</sub> S, ZnS and SnS <sub>2</sub> ”	Composto químico	
“Cu <sub>2</sub> S, ZnS and SnS <sub>2</sub> was cold-pressed at 250 MPa.”	Condições de Pressão	
“After a deposition, thin films were annealed in atmosphere of Ar+S <sub>2</sub> (g).”	Condições atmosféricas	
After a deposition, thin films were annealed in atmosphere of Ar+S <sub>2</sub> (g) at 250–400°C	Temperatura de recozimento	
“X-ray diffractometer (XRD, Phillips) with a CuKα radiation for the phase identification and crystallographic structure.”	Instrumento	
“The surface morphology of the thin films was observed by SEM (Hitachi S-4100).”	Instrumento	
“The Electrical resistance of the thin films was measured using a four-point probe”	Resistência elétrica	
“The Electrical resistance of the thin films was measured using a four-point probe”	Camada absorvente	
“the optical transmittance of the thin films was measured with an UV-VIS-NIR Spectrophotometer”	Transmitância ótica	
“in the wavelength range of 300–2500nm”	Comprimento de onda	
“hν is the photon energy”	Energia do fóton	
“Sputter source”	Fonte de pulverização catódica	
“Radio frequency (13.56MHz)”	Frequência de rádio	

“Purging method”	Método	
“Power”	Potência	
“Sputtering time”	Tempo de pulverização	
“Substrate temperature”	Temperatura do substrato	
“Room temperature”	Condições atmosféricas	
Londhe, P. U., Rohom, A. B., Lakhe, M. G., Bhand, G. R., & Chaure, N. B. (2016). Electrochemically synthesized CuInSe 2 thin films from non-aqueous electrolyte for solar cell applications. Semiconductor Science and Technology 31(12), 125009. <a href="https://doi.org/10.1088/0268-1242/31/12/125009">https://doi.org/10.1088/0268-1242/31/12/125009</a>		
“the cell efficiency is comparatively low (~10%)”	Eficiência	Introdução
“low-cost methods are desirable to prepare the absorber layers”	Técnica de produção de camadas absorventes	
“Dimethyl sulfoxide (DMSO) was used as the solvent”	Solvente	
“onto FTOcoated glass substrate”	Tipo de Substrato	
“from ethylene glycol (EG), which has high viscosity”	Viscosidade do composto	
“from ethylene glycol (EG), which (...) a high boiling point (197 °C)”	Ponto de ebulição do composto	
“onto FTO coated glass substrates”	Substrato	Materials and methods
“onto FTO coated glass substrates at 130 °C”	Temperatura do substrato	
“onto FTO coated glass substrates at 130 °C using Biologic potentiostat/ galvanostat model SP150.”	Instrumento	
“The solution matrix contains copper (II) chloride (CuCl2), indium chloride (InCl3) and selenium tetrachloride (SeCl4)”	Matriz da solução	
“potential range −1.1 to −1.5 V”	Potencial de variação	
“after the deposition, the layers were ultrasonically cleaned in boiled distilled water to remove the residue and/or the traces of EG.”	Tipo de limpeza do substrato	
“Subsequently, the layers were dried in a stream of N2 gas”	Secagem da amostra	
“above the selenium ingots of ~1 gm.”	Tipo de lingote	
“the chamber was heated at ~200 °C”	Temperatura de recozimento	
“for 15 min”	Tempo de recozimento	
“Various characterization techniques are used to study the	Técnica de caracterização	

properties of as-deposited and selenized CIS thin films deposited at various potentials.”		
“The surface topography was examined using a JEOL JSM-6360A scanning electron microscopy (SEM) at accelerating voltage 20 kV and probe current 1 nA”	Instrumento	
“radiation of wavelength 1.5405 Å”	Comprimento de onda	



## Anexo IV. Tabela de resultados do Caso de Uso: Energias Renováveis - Produção de camadas absorventes - Filmes Finos

Input do curador de dados		Input do investigador		
Categorização de descritores	Descritores apresentados ao investigador	Anotação	Comentário	Sugestão de melhoria
Propriedades da amostra	Elemento químico	Copper, Indium, Gallium, selenium or Sulfur		
	Composto químico	CuInGaS thin film, CIS, CZTS, ZnO, CuO, CuFeS		
	Intervalo da banda	For CIGS 1,5 ev		
	Coefficiente de absorção	10 <sup>5</sup>		
	Rendimento do composto	19%		
	Estado físico do composto	Thin film		
	Potencial de deposição	E=0,7V		
	Tipo de semicondutor	Copper indium gallium selenium or sulfur		
	Viscosidade do composto		Not necessary	
	Ponto de ebulição do composto		Not necessary	
	Potencial de variação	0,7v-1,2v		
	Agente complexante	Acid citric		
	Solvente	Distilled water		
	Soluto	CuCL <sub>2</sub> , InCl <sub>3</sub> , H <sub>2</sub> SeO <sub>3</sub> , SC(NH <sub>2</sub> ) <sub>2</sub>		
	Tipo de reação	$\text{Cu}^{2+} + \text{In}^{3+} + 2\text{H}_2\text{SeO}_3 + 13\text{e} + 8\text{H}^+ \rightarrow \text{CuInSe}_2 + 6\text{H}_2\text{O}$		
	Reagentes		No need	
	Composição do banho	CuInS <sub>2</sub> -CuInSe <sub>2</sub>		
	Elétrodo de trabalho	ITO Glass		
	Elétrodo de referência	Ag/AgCl		
	Contra elétrodo	Platine		
	Potencial catódico	-750mV		
	Camada absorvente	CIGS		

	Resistência elétrica		I can't find the exact resistance	
	Transmitância ótica			Transmitância
	Comprimento de onda	400nm-800nm		
	Matriz da solução		No need	
	Energia do fóton		No	
	Fonte de pulverização catódica			
	Frequência de rádio		No	
	Potência da amostra		No	
	Secagem da amostra			
	Tipo de lingote	ITO		
	Tipo de Substrato	ITO-Si-Glass		
	Dimensão do substrato	1,5cm- 1cm		
	Tipo de limpeza do substrato	Ultrason		
<b>Instrumento características e calibrações</b>	Instrumento	Potentiostat		
<b>Tempo</b>	Tempo de secagem da amostra	10 min		
	Tempo de deposição	Between 30 min intel 1 hour		
	Tempo de limpeza	10 min		
	Tempo de estabilização da temperatura			
	Tempo de pulverização		its other technique i don't use	
	Tempo de recozimento	5 min intel 30 min		
<b>Temperatura</b>	Temperatura da atmosfera	T=38°C		
	Temperatura de secagem da amostra	60°C		
	Temperatura de recozimento	370°C		
	Temperatura do substrato	Room temperature		
<b>Método</b>	Método	Electrodeposition, pyrolysis, sol-gel assisted by spin coating, laser ablation, PVD, CVD (Thin-Film		

		Deposition Processes and Technologies)		
<b>Técnica</b>	Técnica de produção de camadas absorventes			
	Técnica de caracterização	XRD, TEM, MEB, UV-visible, four probes method		
<b>Outros</b>	Condições de pressão	1 Atm		
	Condições atmosféricas	Temperature ambien		
	Condições de temperatura	T=38°C		
	Energia do intervalo			
	Índice de refração			
	Coeficiente de extinção			
	Constante dielétrica			Parte real da constante dielétrica  Parte imaginária da constante dielétrica

## Anexo V. Tabela de análise do Caso de Uso da Física - Síntese de nanopartículas

Expressão em linguagem natural que suscita interesse	Sugestão de descritores em linguagem natural	Nome da secção analisada (atribuído pelo autor)
Kazemizadeh, F., Malekfar, R., & Parvin, P. (2017). Pulsed laser ablation synthesis of carbon nanoparticles in vacuum. Journal of Physics and Chemistry of Solids, 104(December 2016), 252–256. <a href="https://doi.org/10.1016/j.jpcs.2017.01.015">https://doi.org/10.1016/j.jpcs.2017.01.015</a>		
“Recently, carbon nanoparticles (CNPs) have attracted an intense attention for their prominent properties. This new class of carbonbased nanostructures is nontoxic, biocompatible, physicochemically and photochemically stable, high aqueous soluble, and shows strong, size dependent and non-blinking optical luminescence emission”	Amostra	Introduction
“Methods for CNPs synthesis generally are classified into two main chemical and physical categories.”	Método de síntese da amostra	
“since the synthesis process usually does not require expensive equipment.”	Instrumento de síntese da amostra	
“Nevertheless, all effort to control the size of particles by chemical methods is restricted to nucleation and growth limitation”	Dimensões da amostra	
“Moreover, PLA in liquid is considered as a cold temperature method and does not take the advantages of synthesis in high temperatures.”	Temperatura de síntese	
“In fact temperature leads to a crystalline structure that is one of the affecting factor on band gap and photoluminescence.”	Intervalo da banda	
“On the other hand, in PLA in vacuum technique, carrier gas pressure and the laser parameters (e.g. pulse width) are perfect tools to control the size of particles”	Gás carreador	
“On the other hand, in PLA in vacuum technique, carrier gas pressure and the laser parameters (e.g. pulse width) are perfect tools to control the size of particles”	Pressão do gás carreador	

“On the other hand, in PLA in vacuum technique, carrier gas pressure and the laser parameters (e.g. pulse width) are perfect tools to control the size of particles”	Largura do pulso do laser	
“The target was embedded in a quartz tube, 5 cm in diameter and 80 cm in length”	Revestimento da amostra	Material and methods
“The target was embedded in a quartz tube, 5 cm in diameter and 80 cm in length”	Dimensões do revestimento	
“The tube was put in a furnace, 900 °C”	Temperatura de síntese	
“and pressurized at 300 mbar”	Pressurização	
“Fundamental mode of a Q-switched pulsed Nd-YAG laser”	Instrumento de emissão de radiação	
“Fundamental mode of a Q-switched pulsed Nd-YAG laser (1064 nm in wavelength)”	Comprimento de onda do pulso	
“with 10 ns pulse duration”	Tempo de duração do pulso	
“and repetition rate of 5 Hz”	Frequência do pulso	
“The spot area of laser beam on the target was 15 mm <sup>2</sup> ”	Área de incidência do feixe	
“its fluence was 1.3 J/cm <sup>2</sup> .”	Influência do feixe	
“Carbon atoms generated by ablation were gathered on a collector at a water-cooled metal surface in the end of tube.”	Coletor	
“Acid treatment and surface passivation were carried out using a standard typical process”	Método de tratamento e passivação da amostra	
“Dispersed CNPs in ethanol (Fig. 2a) were refluxed in nitric acid (2.0 M) for 5 h”	Concentração de refluxante	
“left to cool down to room temperature”	Condições atmosféricas	
“Organic molecule used here for surface passivation was amino-terminated PEG <sub>1000N</sub> ”	Molécula de passivação da superfície	
“15 mg amine-terminated PEG <sub>1000N</sub> was added to 7 mL of acid treatment CNP solution.”	Solução	
“15 mg amine-terminated PEG <sub>1000N</sub> was added to 7 mL of acid treatment CNP solution.”	Concentração da molécula de passivação	
“215 mg amine-terminated PEG <sub>1000N</sub> was added to 7 mL of acid treatment CNP solution.”	Concentração da solução	

“The mixture was heated at 120 °C”	Temperatura de aquecimento da mistura	
“The mixture was heated at 120 °C under nitrogen over 96 h.”	Tempo de aquecimento da mistura	
“In order to remove PEG <sub>1000N</sub> from CNPs, we used two step centrifuging process in 4000 and 6000 rpm for 30 min.”	Método de remoção da molécula de passivação	
“Characterization techniques employed on the products include: FESEM analysis”	Técnica de caracterização	
“Characterization techniques employed on the products include: FESEM analysis that was conducted using TE-SCAN-MIRA3”	Instrumento de análise FESEM	
“XRD pattern was obtained using a X’ Pert MPD (Philips) Xray diffractometer with Co K $\alpha$ radiation ( $\lambda=1.79 \text{ \AA}$ )”	Instrumento de medição de Difração de Raios X	
“Raman spectroscopy was recorded using Nicolet Almega dispersive Raman spectrometer ( $\lambda_{\text{ex}}=532 \text{ nm}$ )”	Instrumento de registo da espectroscopia	
“Moreover, UV/VIS absorption and FTIR was performed using UV - Vis - NIR spectrophotometer (PG instruments model T80+) and PL spectra by UV/Vis spectrophotometer) ÚNICO Model 4802).”	Instrumento de absorção UV/ VIS	
Mostafa, A. M., Yousef, S. A., Eisa, W. H., Ewaida, M. A., & Al-Ashkar, E. A. (2017). Synthesis of cadmium oxide nanoparticles by pulsed laser ablation in liquid environment. <i>Optik</i> , 144, 679–684. <a href="https://doi.org/10.1016/j.ijleo.2017.06.065">https://doi.org/10.1016/j.ijleo.2017.06.065</a>		
“when the size selected of semiconductors is less than 10 nm”	Dimensões do semicondutor	Introduction
“Therefore different types of techniques are focused on this scale. Wet chemical methods, one of these techniques, like chemical precipitation method are the best one to produce this scale”	Método de síntese da amostra	
“In these methods, the reducing agent is used alot to complete a reduction step”	Agente redutor	
“In this method, nanomaterials could be produced with or without the presence of stabilizer.”	Estabilizador	

“Cadmium oxide, CdO, is an important II–VI n-type semiconductor”	Semicondutor	
“X-Ray diffraction (XRD) is used to determine the particle size and crystalline phase.”	Instrumento de medição do tamanho da partícula	
“X-Ray diffraction (XRD) is used to determine the particle size and crystalline phase.”	Tamanho da partícula	
“The formation of rosary shape in nanoscale is confirmed by high resolution transmission electron microscope (HRTEM).”	Instrumento de análise da forma	
“The function groups of produced nanomaterials and its optical properties is studied using spectroscopic device of ultraviolet – visible – near infrared spectrophotometer (UV-VIS-NIR).”	Instrumento de análise das propriedades óticas	
” Cd metal granulated sheets ( $\geq 99.9\%$ )”	Amostra	Experimental section
“Cd metal granulated sheets ( $\geq 99.9\%$ ) were purchased from BDH chemical Ltd pool, England.”	Produtor da amostra	
“All reagents were analytical grade and used without further purification”	Reagente	
“Water was purified by a Millipore Ultrapure water system”	Instrumento de purificação da água	
“Preparation of colloids by PLAL”	Técnica de preparação de colóides	
“Cleaned Cd sheets were immersed in 10 ml of deionized water.”	Solução	
“Cleaned Cd sheets were immersed in 10 ml of deionized water.”	Concentração do solvente	
“The target was irradiated with the second harmonic of apulsed Nd:YAG laser”	Instrumento de irradiação de harmónico	
“apulsed Nd:YAG laser operating at 1064 nm”	Comprimento da irradiação	
“with pulse repetition rate 10 Hz”	Frequência do pulso	
“pulse width 7 ns”	Tempo de duração do pulso	
“and laser energy 80 mJ/pulse”	Energia do laser por pulso	
“The laser beam was focused on the target surface using a lens of focal length 70 mm.”	Distância focal da lente	
“The ablation was carried out for 10 min.”	Tempo de ablação	

“The depth of the solution layer above the target was about10 mm.”	Profundidade de solução acima da amostra	
“During that process, the solution was stirred by rotationalmoving to prevent shielding from produced nanoparticles”	Agitação da solução	
“Optical properties of samples were measured using a JASCO 570 spectrophotometer, Japan.”	Instrumento de análise das propriedades óticas	
“The crystalline structure ofsamples was characterized using XRD (Schimadzu 7000, Japan)”	Instrumento de caracterização da estrutura cristalina	
“operating with Cu K $\alpha$ radiation”	Radiação	
“(λ = 0.154060 nm)”	Comprimento de onda	
“with scanning rate of 4° min <sup>-1</sup> for 2θ values between 10 and 80°”	Taxa de varredura	
“left to dry at roomtemperature”	Temperatura de secagem da amostra	
“left to dry at roomtemperature for 12 h”	Tempo de secagem da amostra	
“while the excess solution was wicked away by a filter paper.”	Instrumento de remoção do excesso de solução	
“EDX measurements were also performedon single particles for chemical analysis using scanning electron microscopy on a Quanta FEG 250 electron microscope.”	Instrumento de análise química	
Valverde-Alva, M. A., García-Fernández, T., Villagrán-Muniz, M., Sánchez-Aké, C., Castañeda-Guzmán, R., Esparza-Alegría, E., ... Herrera, C. E. M. (2015). Synthesis of silver nanoparticles by laser ablation in ethanol: A pulsed photoacoustic study. Applied Surface Science, 355, 341–349. <a href="https://doi.org/10.1016/j.apsusc.2015.07.133">https://doi.org/10.1016/j.apsusc.2015.07.133</a>		
“Until now, different methods, such as evaporation/condensation techniques, chemical and photochemical reduction, tollens and biological synthesis and laser ablation in liquids, have been used to fabricate Ag-NPs”	Método de síntese da amostra	Introduction
“These waves can be detected outwardly as acoustic waves by using, for example, a piezoelectric transducer.”	Instrumento de detecção de vibrações	



“We try to prove that it is possible to use the pulsed PA technique for determining in situ the productivity per pulse”	Produtividade por pulso	Experimental
“and concentration of the Ag-NPs produced by laser ablation in ethanol.”	Concentração da amostra	
“The system uses a pulsed Nd:YAG laser model Surelite III from Continuum”	Instrumento de emissão de radiação	
“with 7 ns of pulse duration”	Tempo de duração de pulso	
“and 1064 nm of wavelength”	Comprimento de onda do pulso	
“As a target, we used a 10 mm × 10 mm and 2 mm in thickness”	Dimensões da amostra	
“silver foil piece 99% pure”	Amostra	
“from Sigma–Aldrich”	Produtor da amostra	
“Ag-NPs were synthesized in a glass cuvette”	Recipiente de síntetização da amostra	
“filled with 10 mL of ethanol”	Solução	
“filled with 10 mL of ethanol”	Quantidade de solução	
“Ag-NPs were synthesized in a glass cuvette filled with 10 mL of ethanol having the following dimensions: 32 mm in height, 18 mm in width and 19 mm in length.”	Dimensões do recipiente de síntetização da amostra	
“A mirror and a spherical convex lens with a focal length of 75 mm”	Distância focal da lente	
“The target was placed 12±9 mm far from the base of the glass cuvette.”	Distância à base do recipiente de síntetização	
“During the ablation process, the target was moved in an X–Y plane parallel to the base of the cuvette to reduce the number of laser pulses irradiating the same spot area”	Agitação da solução	
“For the PA analyses, two piezoelectric sensors”	Instrumento de detecção de vibrações	
“The PA signals measured by both piezoelectric sensors were captured using an oscilloscope of Tektronix model 5054B.”	Instrumento de captura de vibrações	
“The transmitted energy (E <sub>out</sub> ) was measured by means of a calibrated pyroelectric head Vector PHD50 of Scientech”	Instrumento de medição da energia transmitida	
“SEM analyses were performed using a ESEM FEI-Quanta 200 microscope”	Instrumento de análise por microscopia eletrônica de varredura	

<p>“The silver concentration in the colloids was determined by using a Perkin Elmer atomic absorption spectrometer, model 3100, equipped with a silver hollow cathode lamp and a current intensity of 15 mA.”</p>	<p>Instrumento de identificação da concentração da amostra nos coloides</p>	
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------	--

## Anexo VI. Tabela de resultados do Caso de Uso da Física - Síntese de nanopartículas

Input do curador de dados		Input do investigador		
Categorização de descritores	Descritores apresentados ao investigador	Anotação	Comentário	Sugestão de melhoria
Propriedades da amostra	Amostra	GSG5 - nanopartículas de Gd <sub>5</sub> Si <sub>1.8</sub> Ge <sub>2.2</sub>		
	Produtor da amostra	Eu		
	Concentração da amostra	0.5mg		Massa da amostra é mais apropriado para este caso, mas Concentração da amostra também é importante para outro tipo de experiências. E também pode ser importante na minha experiência, nos compósitos.
	Dimensões da amostra			
	Gás carreador			
	Pressão do gás carreador			
	Revestimento da amostra	Substrato silício	Revestimento não é o melhor termo, mas muitas pessoas usam-no.	
	Dimensões do revestimento	1x2mm		
	Pressurização			
	Concentração de refluxante			

	Molécula de passivação da superfície			
	Concentração da molécula de passivação		Não sabe, não reconhece o termo.	
	Solução		Não usa, afirma ser mais utilizado na química.	
	Profundidade de solução acima da amostra			
	Distância à base do recipiente de síntetização			
	Agitação da solução			
	Concentração da solução			
	Semicondutor			
	Dimensões do semicondutor			
	Agente redutor			
	Estabilizador	Atmosfera de argônio		
	Tamanho da partícula	30 nm		
	Reagente			
	Resistência Milli-Q		Resistência Milli-Q é um tipo de água, reconhece que existe água Milli-Q, mas resistência Milli-Q não sabe o que significa, importava saber se esse tipo de água é usado ou não.	
	Concentração do solvente			
	Quantidade de solução			
	Instrumento de síntese da amostra	pulsed laser deposition		

<b>Instrumento características e calibrações</b>	Recipiente de sintetização da amostra			
	Dimensões do recipiente de sintetização da amostra			
	Instrumento de emissão de radiação	KrF Excimer		
	Largura do pulso do laser	248 nm (comprimento de onda); 500 mJ (energia do pulso); 10 Hz (frequência do pulso); 20 ns (duração do pulso)		Sugere que o termo poderia ser Energia do pulso.
	Comprimento de onda do pulso			
	Comprimento da irradiação			
	Distância focal da lente			
	Frequência do pulso			
	Energia do laser por pulso			Deveria ser antes Energia do pulso o mesmo sugerido antes.
	Instrumento de análise FESEM			
	Instrumento de medição de Difração de Raios X			
	Instrumento de registro da espectroscopia			
	Instrumento de absorção UV/ VIS			
	Instrumento de medição do tamanho da partícula			Os dois descritores repetem a mesma ideia, sugere que

	Instrumento de análise da forma			deveria ser Instrumento de análise da morfologia
	Instrumento de análise das propriedades óticas		Faz sentido para quem trabalha com as questões das propriedades óticas, mas a investigadora não usa.	
	Instrumento de purificação da água		Está relacionado com o uso da água Milli-Q, faz sentido	
	Instrumento de irradiação de harmónico			
	Instrumento de caracterização da estrutura cristalina			
	Radiação			
	Comprimento de onda			
	Taxa de varredura			
	Instrumento de remoção do excesso de solução			
	Instrumento de análise química			
	Instrumento de detecção de vibrações			
	Instrumento de captura de vibrações			
	Instrumento de medição da energia transmitida			
	Instrumento de análise por microscopia eletrónica de varredura		Este descritor é desnecessário uma vez que o descritor Instrumento de análise da morfologia sugerido	

			anteriormente serve para captar esta informação	
	Instrumento de identificação da concentração da amostra nos coloides		É mais utilizado na química	
<b>Tempo</b>	Tempo de duração do pulso	20 ns		
	Tempo de aquecimento da mistura	2h		Sugere tempo de deposição
	Tempo de ablação			
	Tempo de secagem da amostra			
<b>Temperatura</b>	Temperatura de síntese	Temperatura ambiente		
	Temperatura de aquecimento da mistura	2°C		
	Temperatura de secagem da amostra			
<b>Método</b>	Método de síntese da amostra	Gas phase synthesis		
	Método de tratamento e passivação da amostra			
	Método de remoção da molécula de passivação			
<b>Técnica</b>	Técnica de caracterização	HR-TEM, XRD e SQUID		
	Técnica de preparação de coloides			
<b>Outros</b>	Condições atmosféricas	Argon – 1 Torr		
	Produtividade por pulso			
	Intervalo da banda			

	Área de incisão do feixe			
	Influência do feixe			
	Coletor	Substrato de Silício		



## Anexo VII. Tabela de análise do Caso de Uso Engenharia Mecânica - Mecânica de fratura: Feixe Cantiléver

Expressão em linguagem natural que suscita interesse	Sugestão de descritores em linguagem natural	Nome da secção analisada (atribuído pelo autor)
Moberg, A., Budzik, M. K., & Jensen, H. M. (2017). Crack front morphology near the free edges in double and single cantilever beam fracture experiments. <i>Engineering Fracture Mechanics</i> , 175, 219-234. <a href="https://doi.org/10.1016/j.engfracmech.2017.01.030">https://doi.org/10.1016/j.engfracmech.2017.01.030</a>		
“Single [1], and double cantiléver beam tests [2], referred to as SCB and DCB respectively, are the most common experimental procedures used to assess the interfacial fracture toughness of such materials.”	Teste de avaliação da resistência à fratura	Introduction
“a crack length”	Comprimento da fissura	
“ $E_{adh}$ Young’s modulus of the adhesive”	Módulo de Young do adesivo	
“ $F_{ext}$ applied external force”	Força externa aplicada	
“ $G_c$ fracture energy”	Energia da fratura	
“ $G_m$ Energy Release Rate”	Taxa de libertação de energia	
“ $G_e$ calculation error of the energy release rate, $G_e =  G - G_c $ ”	Erro de cálculo da taxa de libertação de energia	
“h thickness of the layer/specimen”	Altura da amostra	
“ $K_m$ stress intensity factor”	Fator de intensidade de stress	
“R radius of longitudinal curvature”	Raio da curvatura longitudinal	
“ $R_t$ radius of transverse curvature”	Raio de curvatura transversal	
“T, $T_i$ shear forces tangent to the crack front”	Força de cisalhamento tangente à frente da fratura	
“W width of the specimen”	Largura da amostra	
“ $I_n$ shear modulus”	Módulo de cisalhamento	
“ $r_{ij}$ stress tensor, $i, j = x, y$ or $z$ for $i \neq j$ ”	Tensor de tensão	
“compact tension (CT) specimens”	Amostra / Espécimen	
“residual stresses”	Tensão residual	
“Poisson’s ratio”	Coeficiente de Poisson	
“1.6 mm thickness”	Altura da amostra/ espécimen	Experimental tests
“25 mm width”	Largura da amostra / espécimen	
“were bonded to form the DCB specimen”	Amostra / Espécimen	
“were bonded to form the DCB specimen using a commercial epoxy, rigid-brittle, adhesive (later referred as $E_p$ )”	Adesivo da amostra	

“Loctite Power Epoxy Universal supplied by Henkel Norden AB, Germany”	Fornecedor do adesivo	
“Once the joint was constituted (24 h curing cycle”	Tempo de cura do adesivo	
“24 h curing cycle in ambient conditions”	Condições atmosféricas	
“adhesive layer thickness of $\approx 0.7$ mm”	Altura da camada adesiva	
“applying a displacement d (= 5 mm)”	Deslocamento da fenda	
“it was placed under the Keyence VR-3200 3D scanning macro/microscope (Keyence, Japan)”	Instrumento de medição 3D	
“For qualitative comparison the SCB specimen were prepared”	Amostra de referência	
“polymethamethacrylate- PMMA (1800 Acrylic supplied by RTP”	Adesivo da amostra de referência	
“polymethamethacrylate - PMMA (1800 Acrylic supplied by RTP co., Winona, USA)”	Fornecedor do adesivo de referência	
”25 mm width”	Largura da acamada adesiva	
“propagation is recognized as negligible compared to the observation time)”	Tempo de observação	
“E (GPa)”	Módulo de Young/ elasticidade	
Zhang, D., Shi, H., Zhu, J., Su, M., & Jin, W. L. (2018). Cover separation of CFRP strengthened beam-type cantilevers with steel bolt anchorage. Engineering Structures, 156(September 2017), 224–234. <a href="https://doi.org/10.1016/j.engstruct.2017.10.056">https://doi.org/10.1016/j.engstruct.2017.10.056</a>		
“previous literature has shown that the application of FRP composites can lead to brittle failure”	Tipo de falha	Introduction
“anchors provides a significant increase in anchorage strength,”	Força de ancoragem	
“anchors provides a significant increase in anchorage strength, as well as ductility enhancement.”	Ductilidade	
Although FRP anchor has its advantage over conventional steel anchor	Ancora	
“17 beam-shear experiments with mid-span precutting fissures”	Tensão	
“The cross-section of the RC beam-type cantilêver specimens”	Amostra/ Espécimen	Experiment program
“RC beam-type cantilêver specimens is 120mm×150 mm, and the length is 600 mm.”	Dimensões da amostra	

“All specimens had a clear span of 400 mm, and employed two deformed bars of diameter 10mm as respective compression and tension reinforcement. 10mm diameter plain bar stirrups spaced at 100mm were placed along the entire length of the beam.”	Compressão da amostra	
“CFRP U-anchor with height of 120mm and width of 200mm were used in another side to guarantee the targeted failure mode.”	Dimensões da ancora	
“the horizontal distance of steel bolt center to the right edge of inner artificial crack (or compression end of cantilever)”	Distância entre a ancora e a extremidade de compressão	
“and the number of steel bolt anchors”	Número de ancoras	
“compressive strength ( $f_{cu}$ ) of three 150mm concrete cubes was 31.27 MPa after 28 days of curing.”	Resistência à compressão	
“The modulus of elasticity ( $E$ ), yield strength ( $f_y$ ) and ultimate strength ( $f_u$ ) of deformed bars are 204.40 GPa, 463.0 MPa and 575.5 MPa”	Módulo de Young/ elasticidade	
“The modulus of elasticity ( $E$ ), yield strength ( $f_y$ ) and ultimate strength ( $f_u$ ) of deformed bars are 204.40 GPa, 463.0 MPa and 575.5 MPa”	Força de rendimento	
“The modulus of elasticity ( $E$ ), yield strength ( $f_y$ ) and ultimate strength ( $f_u$ ) of deformed bars are 204.40 GPa, 463.0 MPa and 575.5 MPa”	Resistência final	
“Based on respective uniaxial tension tests.”	Teste de tensão e deformação	
“Two strips of CFRP laminates were”	Laminado	
“Each layer had a width of 40 mm, length of 330mm and a dry fibre thickness of 0.111 mm.”	Dimensões do laminado	
“The elastic modulus and tensile strength of the CFRP laminate were 241 GPa and 3696 MPa”	Módulo de Young/elasticidade do laminado	
“The elastic modulus and tensile strength of the CFRP laminate were 241 GPa and 3696 MPa”	Resistência à tração do laminado	

Adhesive used for both CFRP laminate bonding and steel bolt anchorage bonding was a thixotropic epoxy resin	Adesivo	
Adhesive used for both CFRP laminate bonding and steel bolt anchorage bonding was a thixotropic epoxy resin with elastic modulus and tensile strength being equal to 2.83 GPa and 52 MPa, respectively.	Módulo de Young/elasticidade do adesivo	
Adhesive used for both CFRP laminate bonding and steel bolt anchorage bonding was a thixotropic epoxy resin with elastic modulus and tensile strength being equal to 2.83 GPa and 52 MPa, respectively.	Resistência à tração do adesivo	
“The following procedure was adopted for specimen preparation”	Método de preparação da amostra	
“All the specimens were then placed in a room-temperature environment”	Condições atmosféricas	
“All the specimens were then placed in a room-temperature environment for seven days for the adhesive curing before testing.”	Tempo de cura do adesivo	
“Two strain gauges were glued to the CFRP laminate (Fig. 1) to measure its strain response.”	Instrumento de medição de deformação	
“six strain gauges were attached to the tension steel bars with three strain gauges to each bar”	Instrumento de medição de tensão	
“The deflection was measured by means of LVDTs placed at the loading point and supports of the specimens, respectively”	Instrumento de medição da deflexão	
“Specimens were tested monotonically until failure by a load controlled hydraulic jack at a constant loading rate of 2 kN/min.”	Instrumento de teste de falha	
Wu, N., & Wang, Q. (2011). Experimental studies on damage detection of beam structures with wavelet transform. International Journal of Engineering Science, 49(3), 253–261. <a href="https://doi.org/10.1016/j.ijengsci.2010.12.004">https://doi.org/10.1016/j.ijengsci.2010.12.004</a>		
“Damage detection techniques were provided by researchers, including the crack detection using lamb wave actuated by piezoelectric actuator”	Técnica de detecção de dano	Introduction

“developed another experimental method using digital camera to get the deflection field of a cracked cantilever beam under bending”	Instrumento de captura do campo de deflexão	Experimental set up
“The damage detection effects were obvious from their findings when the crack depth was larger than 50% of the thickness of the cracked beam.”	Profundidade da fratura	
“In the following experimental studies of the crack detection using the spatial Gabor wavelet transform, a high resolution laser profile sensor (CrossCheck laser profile sensor model CC3100-30) is employed to measure the deflection profile”	Instrumento de medição da deflexão	
“Thus, the valid width measurement range will be around 3.0 cm”	Faixa de medição válida	
“of a cracked cantilever aluminum beam subjected to a static displacement at its free end.”	Amostra / Espécimen	
“The length and thickness of the cantilever aluminum beam are 55 cm and 0.3 cm, respectively”	Comprimento da amostra	
“The length and thickness of the cantilever aluminum beam are 55 cm and 0.3 cm, respectively”	Altura da amostra	
“A transverse displacement of 3.5 cm is applied at its free end”	Deslocamento transversal	
“The crack with the width of 0.1 cm”	Largura da fissura	
“The crack (...) is located on the lower surface of the b-eam structure”	Localização da fissura	

## Anexo VIII. Tabela de resultados do Caso de Uso Engenharia Mecânica - Mecânica de fratura - Feixe Cantiléver

Input do curador de dados		Input de comparação com curador de dados	
Categorização de descritores	Descritores apresentados ao curador de dados	Comentário	
Propriedades da amostra	Amostra/ Espécimen	Neste domínio há uma preferência por espécimen, sendo a proposta do curador Espécimen	O curador identificou a necessidade de definir um descritor capaz de captar informação relevante sobre as propriedades da amostra, assim define <i>Propriedades da amostra</i> , este descritor permite que o investigador possa anotar livremente o que considerar relevante.
	Dimensões da amostra		
	Altura da amostra	O descritor apresenta-se na proposta do curador	
	Comprimento da amostra	O descritor apresenta-se na proposta do curador	
	Largura da amostra	O descritor apresenta-se na proposta do curador	
	Adesivo da amostra		
	Amostra de referência		
	Adesivo da amostra de referência		
	Fornecedor do adesivo de referência		
	Largura da acamada adesiva		
	Módulo de Young/ elasticidade		
	Compressão da amostra		
	Taxa de libertação de energia		
	Erro de cálculo da taxa de libertação de energia		
	Força de rendimento		
	Resistência final		
	Resistência à compressão		
	Coeficiente de Poisson		
	Laminado		
	Dimensões do laminado		
	Módulo de Young/elasticidade do laminado		

	Resistência à tração do laminado	
	Adesivo	
	Módulo de Young/ elasticidade do adesivo	
	Resistência à tração do adesivo	
	Número de ancoras	
	Dimensões da ancora	
	Ancora	
	Força de ancoragem	
	Fornecedor do adesivo	
	Altura da camada adesiva	
	Módulo de cisalhamento	
	Força de cisalhamento tangente à frente da fratura	
<b>Instrumento características e calibrações</b>	Instrumento de medição 3D	
	Instrumento de medição de deformação	
	Instrumento de medição de tensão	
	Instrumento de medição de deflexão	O curador definiu apenas o descritor <i>Instrumento</i> .
	Instrumento de teste de falha	
	Instrumento de captura do campo de deflexão	
	Tensor de tensão	
<b>Tempo</b>	Tempo de observação	
	Tempo de cura do adesivo	
<b>Método</b>	Método de preparação da amostra	O curador especificou apenas <i>Método</i>
<b>Técnica</b>	Técnica de detecção de dano	
<b>Outros</b>	Teste de avaliação da resistência à fratura	Identificou um descritor para <i>Velocidade do teste</i> de modo a captar a velocidade a que o teste pressiona o espécimen.
	Teste de tensão de deformação	
	Comprimento da fissura	Capturou o descritor <i>Comprimento da fissura antes da experiência</i>
	Força externa aplicada	
	Energia da fratura	
	Fator de intensidade de stress	
	Raio da curvatura longitudinal	

	Raio de curvatura transversal	
	Tensão residual	
	Condições atmosféricas	O curador identificou <i>Temperatura e Humidade</i>
	Deslocamento da fenda	
	Tipo de falha	
	Ductilidade	
	Tensão	
	Distância entre a ancora e a extremidade de compressão	
	Profundidade da fratura	
	Faixa de medição válida	
	Deslocamento transversal	
	Largura da fissura	
	Localização da fissura	



## Anexo IX. Tabela de análise do Caso de Uso Engenharia Mecânica - Processo de fabrico - Moldagem de chapa

Expressão em linguagem natural que suscita interesse	Sugestão de descritores em linguagem natural	Nome da secção analisada (atribuído pelo autor)
Ke, J., Liu, Y., Zhu, H., & Zhang, Z. (2018). Formability of sheet metal flowing through drawbead—an experimental investigation. Journal of Materials Processing Technology, 254(November 2017), 283–293. <a href="https://doi.org/10.1016/j.jmatprotec.2017.11.051">https://doi.org/10.1016/j.jmatprotec.2017.11.051</a>		
“percentage of elongation is 29%.”	Porcentagem de alongamento	Introduction
“straight or proportional strain path”	Trajectoria de deformação do espécimen	
“absence of bending”	Flexão do espécimen	
“no normal stress or through-thickness stress”	Tensão do espécimen	
“without through-thickness shear”	Tensão de cisalhamento	
“The most commonly used method in experiment determination of the FLD is Nakazima test (out-of-plane) and Marciniak test (in-plane), acquiring strains from uniaxial tensile test to plane strain and then to biaxial tension”	Método de determinação do diagrama de limite de formação	
“by testing specimens with different widths”	Largura do espécimen	
“compared the in-plane stretching”	Alongamento do espécimen no plano	
“with the out-of-plane stretching”	Alongamento do espécimen fora do plano	
“especially when the punch radius was below 10 mm”	Raio de perfuração	
“In the second part, specimens were pulled through various drawbead inserts several times, then cut to the uniaxial tensile test specimens and tested by the conventional uniaxial tensile test”	Teste de referência	Experiment procedures
“conventional uniaxial tensile test (as a reference)”	Teste de referência	
“drawbead test (nine groups of drawbead inserts, passing through one time, two times and three times)”	Teste para a caracterização mecânica	

“Chemical composition of WDQ steel by weight percentages”	Composição química do espécimen	
“Yield strength $\sigma_y$ (Mpa)”	Força de rendimento do espécimen	
“Tensile strength $\sigma_u$ (Mpa)”	Resistência à tração do espécimen	
“Anisotropy index r”	Índice de anisotropia do espécimen	
“Hardening exponent n”	Expoente endurecedor do espécimen	
“Hardening coefficient K”	Coeficiente de endurecimento	
“Young’s modulus E(Gpa)”	Módulo de Young ou módulo de elasticidade	
“Total Elongation $A_t$ (%)”	Elongação total	
The blank material is WDQ steel of 0.6 mm thickness with good ductility, widely used automotive steel provided by Research and development center of Wuhan iron and steel (group) corp.	Espécimen de referência	
“The mechanical properties of WDQ steel tested on Zwick Z050 electronic tensile testing machine”	Instrumento de teste	
“The maximum normal force of the machine is 14,000 N”	Força máxima do instrumento	
“while the pulling force is 30,000 N”	Força de tração do instrumento	
“We have also tested a few of materials (DC03, DC04) with the thickness of 1.0 mm and 1.2 mm and the results are the same and valid for thicker materials certainly.”	Espessura do espécimen	
“The length of the specimen is designed 250 mm”	Comprimento do espécimen	
“and the width of specimen is designed as 30 mm”	Largura do espécimen	
“as the width of drawbead groove is only 70 mm”	Largura do sulco do pino de tração	
“surface of the specimen using electrochemical corrosion method”	Método	
“Before that, the specimens were degreased and cleaned with alcohol.”	Limpeza do espécimen	
“The Auto Grid Com-smart device was used to take pictures of deformed grids on the surface of the specimen when the specimen has been pulled through drawbead inserts, and the length stain, width strain and	Instrumento de medição da deformação	

thickness strain were measured by software analysis.”		
“Adjusting the clamping unit down to the location about $30 \pm 3$ mm to the upper surface of the black box in”	Unidade de aperto da máquina de teste	
“Setting the pulling speed 2.1 mm/s, and then pulling the specimen through drawbead inserts by driving the hydraulic equipment.”	Velocidade de tração da máquina de teste	
Silva, M. B., Isik, K., Tekkaya, A. E., Atkins, A. G., & Martins, P. A. F. (2016). Fracture toughness and failure limits in sheet metal forming. <i>Journal of Materials Processing Technology</i> , 234, 249–258. <a href="https://doi.org/10.1016/j.jmatprotec.2016.03.029">https://doi.org/10.1016/j.jmatprotec.2016.03.029</a>		
“As a result of this, research has been focused on the formability limits at the onset of plastic instability (also known as the forming limit curves, FLC’s).”	Curva limite de formação	Intoduction
“Experimentation the investigation was carried out in aluminium AA1050-H111”	Espécimen	Experimentation
“aluminium AA1050-H111 sheets with 1 mm thickness”	Espessura do espécimen	
“The mechanical characterization of the aluminium AA1050-H111 sheets at room temperature”	Temperatura de caracterização mecânica do espécimen	
“was carried out by means of tensile tests in specimens cut”	Teste para a caracterização mecânica	
“The methodology followed the ASTM standard E8/E8 M (ASTM E8/E8 M, 2013)”	Método	
“The stress strain curve was approximated by the following Ludwik Hollomon’s equation,”	Equação da curva de tensão-deformação	
“Modulus of elasticity (GPa)”	Modulo de Elasticidade do espécimen	
“Yield strength (MPa)”	Força de rendimento do espécimen	
“Ultimate tensile strength (MPa)”	Resistência máxima de tração do espécimen	
“Elongation at break (%)”	Alongamento na ruptura	
“Anisotropy coefficient”	Coefficiente de anisotropia	
“The formability limit by necking (FLC) was determined”	Curva limite de conformação	
“The formability limit by necking (FLC) was determined by means of sheet formability tests (tensile, Nakazima, bulge and hemispherical dome tests) that covered strain	Teste de definição da curva limite de conformação	

paths from uniaxial to biaxial stretching conditions.”		
“The procedure utilized for determining the in-plane strains ( $\epsilon_1$ , $\epsilon_2$ ) at the onset of necking involved electrochemical etching of a grid of overlapping circles”	Método	
“2 mm initial diameter on the surface of the test specimens”	Diâmetro inicial do espécimen	
“The characterization of the formability limits by fracture (FFL and SFFL) included additional results from double-notched test specimens loaded in tension and in plane torsion.”	Limite de formabilidade por fratura	
“Characterization of fracture toughness $R_I$ (mode I) using double notched test specimens loaded in tension is summarized”	Resistência à fratura	
Lou, Y., Chen, L., Clausmeyer, T., Tekkaya, A. E., & Yoon, J. W. (2017). Modeling of ductile fracture from shear to balanced biaxial tension for sheet metals. International Journal of Solids and Structures, 112, 169–184. <a href="https://doi.org/10.1016/j.ijsolstr.2016.11.034">https://doi.org/10.1016/j.ijsolstr.2016.11.034</a>		
“Failure modes in metal forming are mainly classified into three types: necking, ductile fracture, and wrinkling.”	Condição de Fratura	Introduction
“The drawback of this study is that the specimen dimensions were not given”	Dimensão do espécimen	
“uncertainty of the reported failure strain was not analyzed”	Tensão de falha	
“introduced an alternative experimental methodology to determine the fracture locus by combining a reduced number of incremental sheet forming and in-plane shear tests with appropriate procedures to determine the fracture strains.”	Método	
“It was further extended to couple the effect of strain rate”	Taxa de deformação do espécimen	
“It was further extended to couple the effect of strain rate and temperature on ductile fracture”	Temperatura na fratura dúctil	
The extended criterion is denoted as the DF2014 criterion.	Critério	

“For forming of advanced metals of advanced high strength steels, aluminum alloys, titanium alloys and magnesium alloys, ductile fracture can take place in shear, uniaxial tension, plane strain tension and balanced biaxial tension.”	Condições de Fratura	Experiments
“Fracture strains are measured by the reverse engineering method for AA6082 T6 in shear, uniaxial tension, plane strain tension and balanced biaxial tension.”	Método de medição de fratura	
“The metal used in this study is a high strength aluminum alloy AA6082 T6”	Espécimen	
“Dogbone specimens are cut by a laser cutting machine (Trumpf LASERCELL TLC 1005) along three different directions of rolling direction (RD), diagonal direction (DD) and transverse direction (TD).”	Instrumento de corte	
“The specimens are strained at a constant velocity of 0.1 mm per second”	Velocidade de estiramento do espécimen	
“a Zwick 250 kN Allround universal testing machine.”	Instrumento de estiramento	
“Two extensometers are used to measure the deformation along width and longitudinal directions”	Instrumento de medição da deformação	
“Therefore, the plasticity of the alloy is modeled by the isotropic Drucker yield function for FCC metals”	Função de modelação de plasticidade	
“It is observed that necking occurs at a very low strain of about 0.1.”	Deformação de estreitamento	
“The flow curve after necking is obtained by a reverse engineering method which was also used to obtain the strain hardening of DP980 (t1.2) at large strain after necking”	Método de obtenção da curva de fluxo	
“The adjusted stress-strain curve is then employed to describe the strain hardening behavior of AA6082 T6.”	Comportamento de endurecimento da tensão do espécimen	Experiments
“All these specimens are manufactured by laser cutting along RD since anisotropy is not considered in this study.”	Fabrico do espécimen	

“The velocity of the crosshead is set to 0.3 mm per minute for specimens with a central hole and inplane shear specimens while notched specimens are loaded at a velocity of 0.5 mm per minute.”	Velocidade de cruzada	
“the loading processes are recorded by a 3D GOM ARAMIS camera system for the later analysis of ductile fracture.”	Instrumento de registo do processo de carregamento	
“The spatial resolution is about 28 pixels per mm for these three types of tests carried out in a Zwick 250 kN Allround universal testing machine.”	Instrumento de ensaio universal	
“The Nakajima tests are also conducted on a Zwick BUP10 0 0 hydraulic sheet forming tester to study ductile fracture behavior under the balanced biaxial tension.”	Instrumento de teste Nakajima	
“Specimens are clamped with a force of 250 kN.”	Força de aperto do espécimen	
“The punch with a diameter of 100 mm”	Diâmetro de punção	
“The punch with a diameter of 100 mm draws the blank into the die at a constant velocity of 90 mm per minute.”	Velocidade de punção	
“The diameter of the die is 105 mm”	Diâmetro da matriz	
“and the radius of the die chamfer is 10 mm.”	Raio do chanfro da matriz	
“while punch force is measured by a load cell (DL-Systeme, relative accuracy better than 0.24%) for the verification of the plasticity model.”	Instrumento de medição da força de perfuração	